

STIC-Biotech/ChemLib

103 480

**From:** Scheiner, Laurie  
**Sent:** Tuesday, September 09, 2003 8:38 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** seq. search request

RECEIVED

SEP 10 2003

Please search SEQ ID NO:1 of application S.N. 09/849,729. Please also search antisense, if possible. Thanks! (STIC)

Laurie Scheiner  
Art Unit 1648  
CM1 8E05  
308-1122  
8E12

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

# SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 09-16-03  
Searcher: Beverly C 4994  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

### Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_\_ IG  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
✓ \_\_\_\_\_ Other CGN

**THIS PAGE BLANK (USPTO)**

> s (sentinel virus)  
L3 4 (SENTINEL VIRUS)

=> d l3 1-4 bib ab

L3 ANSWER 1 OF 4 USPATFULL on STN  
AN 2003:237383 USPATFULL  
TI **Sentinel virus II**  
IN Liu, Jen-Kuei, Palo Alto, CA, UNITED STATES  
Lewis, Samantha, San Francisco, CA, UNITED STATES  
Batz, Hans-Georg, Tutzing, GERMANY, FEDERAL REPUBLIC OF  
Ramaswamy, Latha, San Jose, CA, UNITED STATES  
Bohenzky, Roy, Mountain View, CA, UNITED STATES  
Lin, Yu-Huei, Sunnyvale, CA, UNITED STATES  
Montiel, Janine, Fremont, CA, UNITED STATES  
Chen, Benjamin P., Fremont, CA, UNITED STATES  
PI US 2003165540 A1 20030904  
AI US 2001-849729 A1 20010504 (9)  
PRAI US 2000-202271P 20000505 (60)  
DT Utility  
FS APPLICATION  
LREP Kenneth J. Waite, Roche Diagnostics Corporation, 9115 Hague Road, Bldg.  
D, P.O. Box 50457, Indianapolis, IN, 46250-0457  
CLMN Number of Claims: 13  
ECL Exemplary Claim: 1  
DRWN 1 Drawing Page(s)  
LN.CNT 1307  
AB The invention relates to a new virus, designated H101.c33 or  
**Sentinel Virus II (SVII)**. Isolated SVII viruses,  
polynucleotides and proteins from SVII viruses, and antibodies which  
bind SVII virus and SVII viral proteins are provided. The  
polynucleotides, proteins, and antibodies of the invention may be used  
to detect SVII virus or infection by SVII virus in a susceptible  
individual. Additionally, polynucleotides of the invention may be  
inserted into recombinant expression vectors for recombinant production  
of viral proteins.

L3 ANSWER 2 OF 4 USPATFULL on STN  
AN 2001:188386 USPATFULL  
TI Hepatitis virus **sentinel virus I (SVI)**  
IN Liu, Jen-Kuei, Palo Alto, CA, United States  
Bohenzky, Roy A., Mountain View, CA, United States  
Lin, Yu-Huei, Sunnyvale, CA, United States  
Chen, Benjamin P., Fremont, CA, United States  
PI US 2001034018 A1 20011025  
AI US 2000-732665 A1 20001208 (9)  
PRAI US 1999-172696P 19991210 (60)  
DT Utility  
FS APPLICATION  
LREP Kenneth J. Waite, Roche Diagnostics Corporation, 9115 Hague Road, Bldg.  
D, P.O. Box 50457, Indianapolis, IN, 46250-0457  
CLMN Number of Claims: 18  
ECL Exemplary Claim: 1  
DRWN 3 Drawing Page(s)  
LN.CNT 1382  
CAS INDEXING IS AVAILABLE FOR THIS PATENT.  
AB The invention relates to a new group of viruses, designated SVI.  
Isolated SVI viruses, polynucleotides and proteins from SVI viruses, and  
antibodies which bind SVI virus and SVI viral proteins are provided. The  
polynucleotides, proteins, and antibodies of the invention may be used  
to detect SVI virus or infection by SVI virus in a susceptible  
individual. Additionally, polynucleotides of the invention may be  
inserted into recombinant expression vectors recombinant production of  
viral proteins.

**THIS PAGE BLANK (USPTO)**

L3 ANSWER 3 OF 4 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN  
AN 2002-062234 [08] WPIDS  
DNC C2002-017791  
TI A new virus associated with cryptogenic non-A/non-G hepatitis, designated **Sentinel Virus II** is useful to detect and treat **Sentinel Virus II** infection.  
DC B04 D16  
IN BATZ, H; BOHENZKY, R; CHEN, B; LEWIS, S; LIN, Y; LIU, J; MONTIEL, J; RAMASWAMY, L; CHEN, B P  
PA (LIUJ-I) LIU J; (HOFF) HOFFMANN LA ROCHE & CO AG F; (HOFF) ROCHE DIAGNOSTICS GMBH; (BATZ-I) BATZ H; (BOHE-I) BOHENZKY R; (CHEN-I) CHEN B P; (LEWI-I) LEWIS S; (LINY-I) LIN Y; (MONT-I) MONTIEL J; (RAMA-I) RAMASWAMY L  
CYC 96  
PI WO 2001085770 A2 20011115 (200208)\* EN 20p  
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW  
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW  
AU 2001065924 A 20011120 (200219)  
EP 1282692 A2 20030212 (200312) EN  
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR  
BR 2001010576 A 20030401 (200327)  
KR 2003032947 A 20030426 (200354)  
CZ 2002003631 A3 20030716 (200355)  
US 2003165540 A1 20030904 (200359)  
ADT WO 2001085770 A2 WO 2001-EP5029 20010504; AU 2001065924 A AU 2001-65924 20010504; EP 1282692 A2 EP 2001-943319 20010504, WO 2001-EP5029 20010504; BR 2001010576 A BR 2001-10576 20010504, WO 2001-EP5029 20010504; KR 2003032947 A KR 2002-714822 20021105; CZ 2002003631 A3 WO 2001-EP5029 20010504, CZ 2002-3631 20010504; US 2003165540 A1 Provisional US 2000-202271P 20000505, US 2001-849729 20010504  
FDT AU 2001065924 A Based on WO 2001085770; EP 1282692 A2 Based on WO 2001085770; BR 2001010576 A Based on WO 2001085770; CZ 2002003631 A3 Based on WO 2001085770  
PRAI US 2000-202271P 20000505; US 2001-849729 20010504  
AB WO 200185770 A UPAB: 20020204  
NOVELTY - A composition comprising isolated **Sentinel Virus II** (SVII) virus, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (1) an isolated polynucleotide which is the sequence fully defined in the specification or its complement;
- (2) an isolated polynucleotide which hybridizes with the above sequence;
- (3) an isolated polynucleotide which encodes an SVII protein or its fragment;
- (4) an isolated polynucleotide whose complement encodes an SVII protein or its fragment, particularly an antisense polynucleotide;
- (5) a composition, particularly a vaccine composition, comprising an isolated SVII protein or its fragment;
- (6) an expression vector comprising an isolated polynucleotide transcription of which produces an SVII antisense polynucleotide;
- (7) an isolated polyclonal antisera which specifically binds an SVII virus or its protein;
- (8) a monoclonal antibody which binds as SVII virus or its protein;
- (9) detecting SVII virus, comprising contacting a sample with the above antibody and detecting antibody:protein complexes;
- (10) detecting SVII virus, comprising contacting a sample with a probe which selectively hybridizes with an SVII polynucleotide and detecting hybridization.

**THIS PAGE BLANK (USPTO)**



ACTIVITY - Anti-viral. No relevant biological data was provided.

MECHANISM OF ACTION - None given.

USE - The invention is used to detect and treat SVII infection.

Dwg.0/1

L3 ANSWER 4 OF 4 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN  
AN 2001-381643 [40] WPIDS  
DNC C2001-116931  
TI Novel virus, designated **sentinel virus** I, associated  
with cryptogenic, nonA-G hepatitis, and polynucleotides and polypeptides  
of virus useful for detecting SVI virus and/or SVI virus infection.  
DC B04 D16  
IN BOHENZKY, R A; CHEN, B P; LIN, Y; LIU, J  
PA (HOFF) ROCHE DIAGNOSTICS GMBH; (BOHE-I) BOHENZKY R A; (CHEN-I) CHEN B P;  
(LINY-I) LIN Y; (LIUJ-I) LIU J  
CYC 92  
PI WO 2001042299 A2 20010614 (200140)\* EN 62p  
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ  
NL OA PT SD SE SL SZ TR TZ UG ZW  
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES  
FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS  
LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL  
TJ TM TR TT TZ UA UG UZ VN YU ZA ZW  
AU 2001022129 A 20010618 (200161)  
US 2001034018 A1 20011025 (200170)  
BR 2000016289 A 20020813 (200262)  
EP 1240189 A2 20020918 (200269) EN  
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT  
RO SE SI TR  
KR 2002065559 A 20020813 (200309)  
JP 2003516136 W 20030513 (200334) 81p  
CZ 2002001962 A3 20030716 (200355)  
ADT WO 2001042299 A2 WO 2000-IB2011 20001208; AU 2001022129 A AU 2001-22129  
20001208; US 2001034018 A1 Provisional US 1999-172696P 19991210, US  
2000-732665 20001208; BR 2000016289 A BR 2000-16289 20001208, WO  
2000-IB2011 20001208; EP 1240189 A2 EP 2000-985731 20001208, WO  
2000-IB2011 20001208; KR 2002065559 A KR 2002-707427 20020610; JP  
2003516136 W WO 2000-IB2011 20001208, JP 2001-543596 20001208; CZ  
2002001962 A3 WO 2000-IB2011 20001208, CZ 2002-1962 20001208  
FDT AU 2001022129 A Based on WO 2001042299; BR 2000016289 A Based on WO  
2001042299; EP 1240189 A2 Based on WO 2001042299; JP 2003516136 W Based on  
WO 2001042299; CZ 2002001962 A3 Based on WO 2001042299  
PRAI US 1999-172696P 19991210; US 2000-732665 20001208  
AB WO 200142299 A UPAB: 20021007  
NOVELTY - A composition (I), comprising isolated **sentinel**  
**virus** I (SVI) virus, is new.  
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the  
following:  
(1) an isolated polynucleotide (II) which:  
(a) selectively hybridizes with a 3847, 3844, 2499, 2499 or 2496  
nucleotide SVI sequence (S1-S5), all fully defined in the specification;  
(b) is complementary to (a);  
(c) encodes an isolated SVI protein or its fragment; or  
(d) is complementary to an isolated polynucleotide encoding an  
isolated SVI protein or its fragment, the nucleotide sequence of (II) is  
distinct from the genomic sequences of TTV strains SANBAN and TUS01;  
(2) a composition comprising an isolated SVI protein or its fragment  
(III) which is serologically distinct from proteins of TTV strains SANBAN  
and TIS01;  
(3) a vaccine composition comprising (III) and an excipient;  
(4) an expression vector comprising an isolated polynucleotide  
encoding an SVI protein or its fragment which is serologically distinct  
from proteins of TTV strains SANBAN and TUS01;  
(5) an expression vector comprising an isolated polynucleotide, where

**THIS PAGE BLANK (USPTO)**

transcription of the isolated polynucleotide results in the production of an SVI antisense polynucleotide which is not an antisense polynucleotide that forms a duplex with an RNA transcript from TTV strains SANBAN and TUS01;

(6) an isolated polyclonal or monoclonal antibody (IV) which binds (III); and

(7) an antibody (V) that binds to an SVI virus whose genome contains (S1-S5).

ACTIVITY - Antiviral.

No biological data is given.

MECHANISM OF ACTION - Vaccine; antisense gene therapy.

USE - (II), (IV), (V) are useful for detecting SVI virus. The method involves contacting a sample with:

(a) (IV) which binds (III) but does not bind TTV strains SANBAN and TUS01 or its proteins, and detecting (IV)-(I) complex; or

(b) (II) which selectively hybridizes to:

(i) a SVI polynucleotide but does not selectively hybridize with polynucleotides of TTV strains SANBAN and TUS01; or

(ii) a target polynucleotide of a viral genome which comprises (S1-S5), and detecting SVI virus by detecting hybridization of (II)-SVI polynucleotide or detecting hybridization of the probe with the target, respectively; or

(iii) the SVI virus and detecting SVI virus by detecting SVI virus-(V) complexes.

(All claimed). Probes and primers derived from SVI polynucleotide sequences which comprise a sequence of (S1)-(S5) is useful for identifying and isolating new variants of SVI. SVI polynucleotides are useful for detecting SVI virus, producing SVI polypeptides, constructing SVI-based expression/transduction vectors and as antisense oligonucleotides or for construction of antisense SVI vectors. Antisense SVI polynucleotides block expression of SVI proteins and/or SVI viral replication in SVI infected cells, and thus are useful for treating SVI infections. SVI polypeptides are useful in vaccines for preventing SVI infection and for treating SVI infection. (IV) or (V) is useful for detecting and/or identifying SVI virus and may also be useful for isolating viral particles and/or viral proteins.

Dwg.0/3

=> s 13 and 11

L4 0 L3 AND L1

=>

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 22:41:54 ; Search time 2084 Seconds

(without alignments)  
7282.856 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371

Sequence: 1 gatcmgaacgcytsgctc.....ctctaccgagcagatc 371

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

```

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	369	99.5	371	6	AX301834
2	366.6	98.8	372	6	AX301835
3	51.2	13.8	232092	2	AC127800
4	51.2	13.8	259827	2	AC137318
5	51.2	13.8	284938	2	AC091335
6	50.2	13.5	1926	6	AR217867
7	50.2	13.5	1931	6	AR083152
8	49.6	13.4	135924	2	AC115529
9	49.6	13.4	236021	2	AC135485
10	49.6	13.4	278501	2	AC098552
11	49.6	13.4	279148	2	AC123507
12	48	12.9	152564	10	AC127549
13	48	12.9	181931	2	AC115847
14	48	12.9	215119	2	AC098082
15	48	12.9	236539	2	AC124637
16	47.8	12.9	160737	2	AC139612
17	47.6	12.8	161046	14	AF081810
18	47.6	12.8	223508	2	AC102480
19	47.6	12.8	234817	10	AL663048
20	46.8	12.6	53724	2	AC131074
21	46.4	12.5	73476	2	AC101502
22	46.4	12.5	193390	10	AC101527
23	46.2	12.5	110000	2	AC097542
24	46	12.4	229583	10	AL844566
25	45.6	12.3	215097	2	AC097088
26	45.6	12.3	226932	2	AC122083
27	45.2	12.2	211526	2	AC120549
28	45	12.1	390	6	AR135147
29	45	12.1	485	8	AF231800
30	45	12.1	1860	1	AV11AD
31	44.8	12.1	3012	14	AF360120
32	44.8	12.1	120627	2	AC104714
33	44.8	12.1	149303	8	AC087412
34	44.8	12.1	226325	2	AC131744
35	44.8	12.1	253946	2	AC105161
36	44.8	12.1	273260	2	AC123888
37	44.8	12.1	340972	2	AC109045
38	44.2	11.9	217932	2	AC115817
39	44	11.9	198281	10	AC122254
40	44	11.9	210519	10	AC125408
41	44	11.9	234130	2	BX284111
42	44	11.9	241633	2	AC096932
43	43.8	11.8	178348	2	AC132931
44	43.8	11.8	184556	2	AC118709
45	43.6	11.8	146766	2	AC105079

## ALIGNMENTS

```

RESULT 1
AX301834
LOCUS AX301834 371 bp DNA linear PAT 30-NOV-2001
DEFINITION Sequence 1 from Patent WO0185770.
ACCESSION AX301834
VERSION AX301834.1 GI:17382895
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1
AUTHORS Liu,J.K., Lewis,S., Bohenzky,R., Lin,Y.H., Ramaswamy,L.,
Montiel,J., Batz,H.G. and Chen,B.
TITLE Sentinel virus II
JOURNAL Patent: WO 0185770-A 1 15-NOV-2001;

```

FEATURES  
Source  
Location/Qualifiers  
1. 371  
/organism="Sentinel virus II"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:179884"

BASE COUNT  
71 a 107 c 109 g 79 t 5 others

ORIGIN  
Query Match 99.5%; Score 369; DB 6; Length 371;  
Best Local Similarity 100.0%; Pred. No. 1,2e-60;  
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GATMGGAACGCTTSGCTCGTGCATGCAAGAGACGGGWTGAAGCGGACGGAGTTGA 60  
1 GATMGGAACGCTTSGCTCGTGCATGCAAGAGACGGGWTGAAGCGGAGTTGA 60  
CGACGACGACGACATTCGATGAAAGATGGACCGCGACGCTCCTTGGCGGGCGGAGCG 120  
CGACGACGACGACATTCGATGAAAGATGGACCGCGACGCTCCTTGGCGGGCGGAGCG 120  
CGAGAACCAAGACGACGACGACGACGCTTACCGCGGACATCCGTTTCTTCTGAGCG 180  
CGAGAACCAAGACGACGACGACGACGCTTACCGCGGACATCCGTTTCTTCTGAGCG 180  
GGATTGACACCTCCGATTCGATGCTGAGATTCGCTGCGTGCAGACGCTGATTCAGC 240  
GGATTGACACCTCCGATTCGATGCTGAGATTCGCTGCGTGCAGACGCTGATTCAGC 240  
GCCGCTTCGTCGACCTGCTTTGACGCGTCGAAACCGGACGCGCTCGAACATGTCAGTT 300  
GCCGCTTCGTCGACCTGCTTTGACGCGTCGAAACCGGACGCGCTCGAACATGTCAGTT 300  
TCCGCTTCAGAGTGAATTCGACCCCGCTTGACCTTGTCTCTCTCTCTCTAC 360  
TCCGCTTCAGAGTGAATTCGACCCCGCTTGACCTTGTCTCTCTCTCTCTAC 360  
CGACGACGATC 371  
CGACGACGATC 371

RESULT 2  
AX301835/c  
LOCUS  
AX301835  
DEFINITION  
Sequence 2 from Patent WO0185770.  
ACCESSION  
AX301835  
VERSION  
AX301835.1 GI:17382896  
KEYWORDS  
Sentinel virus II  
Source  
Sentinel virus II  
Viruses; unclassified viruses.  
REFERENCE  
1 Liu, J. K., Lewis, S., Bohenzky, R., Lin, Y. H., Ramaswamy, L.,  
Montiel, J., Batz, H. G. and Chen, B.  
TITLE  
Sentinel virus II  
JOURNAL  
Patent: WO 0185770-A 2 15-NOV-2001;  
FEATURES  
Location/Qualifiers  
1. 372  
/organism="Sentinel virus II"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:179884"

BASE COUNT  
79 a 109 c 107 g 71 t 6 others

ORIGIN  
Query Match 98.8%; Score 366.6; DB 6; Length 372;  
Best Local Similarity 99.5%; Pred. No. 3.5e-60;  
Matches 369; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GATMGGAACGCTTSGCTCGTGCATGCAAGAGACGGGWTGAAGCGGACGGAGTTGA 60  
1 GATMGGAACGCTTSGCTCGTGCATGCAAGAGACGGGWTGAAGCGGAGTTGA 60  
GATMGGAACGCTTSGCTCGTGCATGCAAGAGACGGGWTGAAGCGGAGTTGA 312  
CGACGACGACGACATTCGATGAAAGATGGACCGCGACGCTCCTTGGCGGGCGGAGCG 120

DB  
CGACGACGACGACATTCGATGAAAGATGGACCGCGACGCTCCTTGGCGGGCGGAGCG 252

QY  
CGAGAACCAAGACGACGACGACGACGCTTACCGCGGACATCCGTTTCTTCTGAGCG 180

DB  
CGAGAACCAAGACGACGACGACGACGCTTACCGCGGACATCCGTTTCTTCTGAGCG 192

QY  
GGATTGACACCTCCGATTCGATGCTGAGATTCGCTGCGTGCAGACGCTGATTCAGC 240

DB  
GGATTGACACCTCCGATTCGATGCTGAGATTCGCTGCGTGCAGACGCTGATTCAGC 132

QY  
GCCGCTTCGTCGACCTGCTTTGACGCGTCGAAACCGGACGCGCTCGAACATGTCAGTT 300

DB  
GCCGCTTCGTCGACCTGCTTTGACGCGTCGAAACCGGACGCGCTCGAACATGTCAGTT 72

QY  
TCCGCTTCAGAGTGAATTCGACCCCGCTTGACCTTGTCTCTCTCTCTCTAC 360

DB  
TCCGCTTCAGAGTGAATTCGACCCCGCTTGACCTTGTCTCTCTCTCTCTAC 12

QY  
CGACGACGATC 371

DB  
CGACGACGATC 1

RESULT 3  
AC127800/c  
LOCUS  
AC127800  
DEFINITION  
Rattus norvegicus clone CH230-30C5. \*\*\* SEQUENCING IN PROGRESS \*\*\*  
ACCESSION  
AC127800.3 GI:25078136  
VERSION  
HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 232092)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregregios, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,  
Hollins, B., Howells, S., Huylk, S., Hume, J., Idelberg, D., Jolivet, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Liu, J.,  
Kowis, C., Kraft, C. U., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D.,  
Lorenshuwa, L., Louisedge, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., McNeill, T. Z., Meenan, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenan, E., Moore, S.,  
Moses, M., Moseley, A., Miner, G., Minna, E., Montemayor, J., Moore, S.,  
Moran, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwokedi, O., Okunolu, G., Olanunsa, A., Olanunsa, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poldinger, A., Popovic, D., Pritmus, E., Pu, L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,









Qy 151 CTACCG 157  
Db 809 GGACGG 803

RESULT 7  
AR083152/c  
LOCUS AR083152 1931 bp DNA linear PAT 01-SEP-2000  
DEFINITION Sequence 2 from patent US 5976807.  
ACCESSION AR083152  
VERSION AR083152.1 GI:10009942  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1931)  
AUTHORS Horlick, R.A., Damaj, B.B. and Robbings, A.K.  
TITLE Eukaryotic cells stably expressing genes from multiple transfected episomes  
JOURNAL Patent: US 5976807-A 2 02-NOV-1999;  
FEATURES  
Source Location/Qualifiers  
1..1931  
BASE COUNT 217 a 873 c 352 g 489 t  
ORIGIN

Query Match 13.5% Score 50.2; DB 6; Length 1931;  
Best Local Similarity 62.2%; Pred. No. 6.8;  
Matches 79; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 31 GAAGACGCGGTGAAGCGGCGGATTGACGACGACGACATTCGATGAAGATGG 90  
Db 929 GGAGGACGGGAGGACGACGCGGAGACGCGGAGACGACGCGGAGACGCGGGA 870  
Qy 91 GACCGCTGACGCTCTTGGCGGGCGGAGCCGAGACCAAGACGACGAGGACGACGCT 150  
Db 869 GGACGACGAGCGGCGGAGACGCGGAGGAGACGACGACGCGGAGGACGCGGAGGAGGACGA 810  
Qy 151 CTACCG 157  
Db 809 GGACGG 803

RESULT 8  
AC115529/c  
LOCUS AC115529 135924 bp DNA linear HTG 30-AUG-2002  
DEFINITION Rattus norvegicus clone RP31-42022 strain Brown Norway, WORKING  
ACCESSION AC115529  
VERSION AC115529.2 GI:22549758  
KEYWORDS HTG: HTGS PHASE2; HTGS\_DRAFT.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 135924)  
AUTHORS Ahlter, N., Antonellis, A., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Boufield, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Latic, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, O.L., Maduro, V.B., Margulies, E.H., Masello, J., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pasutigian, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N., Scheeler, M.G., Sison, C., Scharf, P., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L., Weathery, K.D., Wisgins, L., Young, A., Zhang, L.-H. and Green, E.D.  
JOURNAL NISC Comparative Sequencing Initiative  
REFERENCE 2 (bases 1 to 135924)  
AUTHORS Green, E.D.  
TITLE Direct Submission  
JOURNAL

JOURNAL  
Submitted (20-MAR-2002) NIH Intramural Sequencing Center, 8717  
Grovemont Circle, Galtersburg, MD 20877, USA  
3 (bases 1 to 135924)  
AUTHORS Green, E.D.  
TITLE Direct Submission  
JOURNAL  
Submitted (30-AUG-2002) NIH Intramural Sequencing Center, 8717  
Grovemont Circle, Galtersburg, MD 20877, USA  
On Aug 30, 2002 this sequence version replaced gi:19551131.  
COMMENT  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc\\_zoo@nhgri.nih.gov](mailto:nisc_zoo@nhgri.nih.gov)  
----- Project Information  
Center project name: cxm  
Center clone name: 042022

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 134458 bases at least Q40  
Consensus quality: 134732 bases at least Q30  
Consensus quality: 135037 bases at least Q20  
Insert size: 134000; agarose-fp  
Insert size: 135124; sum-of-contigs  
Quality coverage: 9.95x in Q20 bases; agarose-fp  
Quality coverage: 9.87x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
1 10522: contig of 10522 bp in length  
\* 10523 10622: gap of unknown length  
\* 10623 40668: contig of 30046 bp in length  
\* 40669 40768: gap of unknown length  
\* 40769 70573: contig of 29805 bp in length  
\* 70574 70673: gap of unknown length  
\* 70674 74333: contig of 3660 bp in length  
\* 74334 74433: gap of unknown length  
\* 74434 78760: contig of 4337 bp in length  
\* 78761 78860: gap of unknown length  
\* 78861 85256: contig of 6266 bp in length  
\* 85257 11549: gap of unknown length  
\* 11550 11649: contig of 2693 bp in length  
\* 11650 113764: contig of 2115 bp in length  
\* 113765 113864: gap of unknown length  
\* 113865 135924: contig of 22060 bp in length.  
Location/Qualifiers  
1..135924  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="Brown Norway"  
/db\_xref="taxon:10116"

FEATURES  
source

[illegible]

Egan, A., Escotte, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Huix, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshewal, L., Loulsegod, H., Lozano, R.J., Lu, X., Ma, J.,  
 Maheswari, M., Mahndarte, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milsogajevic, A., Miner, G., Minda, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munizaga, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwaokwelen, O., Okwuonu, G., Olampunsgoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamkoch, C.,  
 Plapper, F., Poudexter, A., Popovic, D., Prizmu, E., Pu, L.,  
 Pizzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Rellly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scheer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Sma's, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabop, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valae, R., Vera, V., Villasa, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, F.,  
 Williams, G., Willson, R., Wlaczky, R., Woodson, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G., and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 236021)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (17-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 236021)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 20, 2002 this sequence version replaced gi:24080625.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: KBL  
 Center clone name: CH230-22003  
 ----- Summary Statistics

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html),
* NOTE: This is a 'working-draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1      132351 contig of 132351 bp in length
*      132352      132451 gap of unknown length
*      132452      236021 contig of 103570 bp in length.

```

1 (bases 10 to 278501)  
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Allen, C., Allen, H., Alsbrooks, S., Amlin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Balandranakite, D., Barber, M., Barnstead, M., Benamed, F.,  
Blawie, K., Blair, J.P., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cadenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., P. Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebrerger, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hui, K. S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kows, C., Kraft, C. L., Lebow, H., Leyvan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D.,  
Lohrensuew, L., Loulgesed, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindaratne, M., Mahmood, M., Mallory, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
Mamun, B., McLeod, M. P., McNeill, T. Z., Meenan, E.,  
Mamun, B., McLeod, M. P., McNeill, T. Z., Meenan, E.,  
Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L.,  
Nandavathi, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwackalameh, O., Okunolu, G., Olapunnsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.,  
Punzo, M., Quiroz, J., Racilin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, M., Saverly, G., Scheper, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajd, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sores, J.,  
Steinle, M., Strong, R., Sutton, A., Svatsek, A., Tabors, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tringey, A., Trejos, Z., Usmani, K.,  
Valdes, R., Vera, V., Villaana, D., Waldron, L., Walker, B., Wang, D.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., Worley, F.,  
Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, Y., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G., and Gibbs, R. A.

Direct Submissions  
Unpublished  
2 (bases 1 to 278501)

Worley, K.C.  
Direct Submission  
Submitted (24-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 278501)

Rat Genome Sequencing Consortium.  
Direct Submissions  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23268777.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.



```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
WILLIAMS,G., WILLSON,R., WLECZYK,R., WOODEN,H., WORLEY,K.,
WRIGHT,D., WRIGHT,R., WU,J., YAKUB,S., YEN,J., YOON,L., YOON,V.,
YU,F., ZHANG,J., ZHOU,J., ZHOU,X., ZHAO,S., DUMM,D., VON
NIEDERHAUSEN,A., WEISS,R., SMITH,D.R., HOLT,R.A., SMITH,H.O.,
WEINSTECKE,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 279148)
Worley K.C.
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 279148)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21909215.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rac/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GYRK
Center clone name: CH230-18C5
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 203914 bases at least Q40
Consensus quality: 206971 bases at least Q20
Consensus quality: 208889 bases at least Q20
Estimated insert size: 222552; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3512: contig of 3512 bp in length
* 3513
* 3612: gap of unknown length
* 3613
* 271054: contig of 267442 bp in length
* 271055
* 271154: gap of unknown length
* 271155
* 272170: contig of 1016 bp in length
* 272171
* 272270: gap of unknown length
* 272271
* 273543: contig of 1273 bp in length
* 273544
* 273643: gap of unknown length
* 273644
* 273500: contig of 1707 bp in length
* 273541
* 275450: gap of unknown length
* 275451
* 277162: contig of 1712 bp in length
* 277163
* 277262: gap of unknown length
* 279148: contig of 1886 bp in length.
* Location/Qualifiers
* 1..279148
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-18C5"
* 3613..5085
misc_feature

```

```

/misc_feature      /note="wgs_end_extension
                   clone_end:T7"
                   4798..5232
                   /note="clone_boundary
                   clone_end:T7
                   site:EcORI
                   end sequence:BH357791"
/misc_feature      complement(266213..267122)
                   /note="clone_boundary
                   clone_end:Sp6
                   site:EcORI
                   end sequence:BH357791"

BASE COUNT      57483 a 50028 c 49540 g 52690 t 65407 others
ORIGIN
Query Match      13.4%; Score 49.6; DB 2; Length 279148;
Best Local Similarity 62.3%; Pred. No. 5.2;
Matches 76; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

28 GCAGAAAGACGCGGWTGAAGCGGACGGGATTGACGACGACGACATTGCGATGAAGA 87
Db 271715 GGAGGAGAGACGAGAGAAAGAGAGACGAGAGAGAGAGAGACGAGAGCAAGGA 271656

88 TGGGACCGCCYGACGTCCTTGGCGGGGCGGAGCGGAGAACCAAGACGACGAGGCGAGGA 147
Db 271655 GGAGGAGAGACGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 271596

148 CG 149
CG 271595 CG 271594
Db

```

```

RESULT 12
AC127549/c 152564 bp DNA linear ROD 21-FEB-2003
LOCUS AC127549
DEFINITION Mus musculus chromosome 5 clone RP24-510G5, complete sequence.
ACCESSION AC127549
VERSION AC127549.3
KEYWORDS GI:28461099
SOURCE HTG.
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 152564)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152564)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 152564)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 152564)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Feb 21, 2003 this sequence version replaced gi:27734044.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0510G05
-----
Location/Qualifiers

```

[illegible]

COMMENT

TITLE  
JOURNAL

-----  
Cook, A., Cooke, P., DeArelino, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Parr, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
Galaan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazars, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,  
Liu G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Milnova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicoli, R., Norby, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Punthiang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Riaback, M., Riley, R., Rise, C.,  
Rogor, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
Vassiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 21, 2002 this sequence was replaced replacing g1:21431359.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----  
Project Information  
Center project name: L24657  
Center clone name: 245\_J\_15  
-----  
Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 175364 bases at least Q40  
Consensus quality: 177921 bases at least Q30  
Consensus quality: 178974 bases at least Q20  
Insert size: 176000; agarose-fp  
Insert size: 179631; sum-of-contigs  
Quality coverage: 7.8 in Q20 bases; agarose-fp  
Quality coverage: 7.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
1  
35917 35916: contig of 35916 bp in length  
\*  
35917 35916: gap of 100 bp  
\*  
36017 36016: contig of 663 bp in length  
\*  
36680 36679: gap of 100 bp  
\*  
36780 37463: contig of 684 bp in length  
\*  
37464 37463: gap of 100 bp  
\*  
37564 38229: contig of 666 bp in length  
\*  
38230 38329: gap of 100 bp  
\*  
38330 38982: contig of 653 bp in length  
\*  
38983 39082: gap of 100 bp  
\*  
39083 39948: contig of 866 bp in length  
\*  
39949 40048: gap of 100 bp  
\*  
40049 40697: contig of 649 bp in length  
\*  
40698 40797: gap of 100 bp  
\*  
40798 42040: contig of 1243 bp in length  
\*  
42041 42140: gap of 100 bp  
\*  
42141 43329: contig of 1189 bp in length  
\*  
43330 43429: gap of 100 bp  
\*  
43430 45637: contig of 2208 bp in length

```
* 45638 45737: gap of 100 bp
* 45738 47265: contig of 1528 bp in length
* 47266 47365: gap of 100 bp
* 47366 49489: contig of 2134 bp in length
* 49500 49599: gap of 100 bp
* 49600 51762: contig of 2163 bp in length
* 51763 51862: gap of 100 bp
* 51863 54660: contig of 2798 bp in length
* 54661 54760: gap of 100 bp
* 54761 58644: contig of 3884 bp in length
* 58645 58744: gap of 100 bp
* 58745 61857: contig of 3112 bp in length
* 61857 61956: gap of 100 bp
* 61957 65844: contig of 3888 bp in length
* 65845 65944: gap of 100 bp
* 65945 69920: contig of 3976 bp in length
* 69921 70020: gap of 100 bp
* 70021 79605: contig of 9585 bp in length
* 79606 79705: gap of 100 bp
* 79706 93801: contig of 14096 bp in length
* 93802 93901: gap of 100 bp
* 93902 108435: contig of 14534 bp in length
* 108436 108535: gap of 100 bp
* 108536 124107: contig of 15572 bp in length
* 124108 124207: gap of 100 bp
* 124208 149441: contig of 25234 bp in length
* 149442 149541: gap of 100 bp
* 149542 181931: contig of 32390 bp in length.
```

FEATURES  
Source

```
1. 181931
   Location/Qualifiers
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /clone_id="RP24-245J15"
   /clone_lib="RPC1-24 Male Mouse BAC"
1. 35316
   /note="assembly_fragment"
   clone_end:SP6
   vector_side:left"
36017..36679
   /note="assembly_fragment"
36780..37463
   /note="assembly_fragment"
37564..38229
   /note="assembly_fragment"
38330..38982
   /note="assembly_fragment"
39083..39948
   /note="assembly_fragment"
40049..40697
   /note="assembly_fragment"
40798..42040
   /note="assembly_fragment"
42141..43329
   /note="assembly_fragment"
43430..45637
   /note="assembly_fragment"
45738..47265
   /note="assembly_fragment"
47366..49499
   /note="assembly_fragment"
49600..51762
   /note="assembly_fragment"
51863..54660
   /note="assembly_fragment"
54761..58644
   /note="assembly_fragment"
58745..61856
   /note="assembly_fragment"
61957..65844
   /note="assembly_fragment"
65945..69920
   /note="assembly_fragment"
```

```
misc_feature 70021..79605
              /note="assembly_fragment"
misc_feature 79706..93801
              /note="assembly_fragment"
misc_feature 93902..108435
              /note="assembly_fragment"
misc_feature 108536..124107
              /note="assembly_fragment"
misc_feature 124208..149441
```

Query Match: 12.9%; Score 48; DB 2; Length 181931;  
Best Local Similarity 61.5%; Pred. No. 11;  
Matches 75; Conservative 1; Mismatches 46; Indels 0; Gaps 0;

```
QY 28 GCAGAAAGACGCGMTGAAGCGGACCGGATTGACGACGACGACGATTCGATGAAAGA 87
Db 83388 GGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83329
QY 88 TGGGACCGCGAGCGCTCTTGGCGGGCGGACCGCGAACAAGACGACGACGAGAGA 147
Db 83328 GGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGA 83269
QY 148 CG 149
Db 83268 CG 83267
```

RESULT 14  
AC098082/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-129A9, \*\*\* SEQUENCING IN PROGRESS  
AC098082  
VERSION  
AC098082.5 GI:30522728  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
1 (bases 1 to 215119)  
Muzny,D.,Marte,, Metzker,M.,Lee,, Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alibrooks,S., Amin,A., Angiano,D.,  
Anyalebechi,T., Aoyagi,A., Ayodeji,M., Bacca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denon,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,  
Gumaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlik,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
Hollins,B., Howells,S., Huily,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kovis,C., Kraft,C.L., Lebow,H., Lervan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,Y., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorenzowa,L., Louised,H., Lozano,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangun,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,  
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nait,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwokeneme,O., Okwuonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,



Pastermak,S., Paul,H., Perez,A., Perez,I., Pfamkoch,C.,  
 Plomper,F., Poldexeter,A., Popovic,D., Primus,E., Pu,L.,L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Reiter,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sander,M., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleciwyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,D., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 215119)  
 Worley,K.C.  
 Direct Submission  
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 215119)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:12855414.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GGFY  
 Center clone name: CH230-129A9  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 201734 bases at least Q40  
 Consensus quality: 203906 bases at least Q30  
 Consensus quality: 205336 bases at least Q20  
 Estimated insert size: 213565; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_diff\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_diff_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 212205: contig of 212205 bp in length  
 \* 212206 212305: gap of unknown length  
 \* 212306 213509: contig of 1204 bp in length

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 21:57:24 ; Search time 221 Seconds  
(without alignments)  
4531.636 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371  
Sequence: 1 gatcmgmgaaacgyttsgrtc.....ctctctaacccgacgacgarc 371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_15Jun03:\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: \*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: \*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: \*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: \*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: \*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: \*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: \*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: \*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: \*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: \*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: \*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: \*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: \*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: \*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: \*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: \*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: \*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: \*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: \*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: \*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: \*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: \*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: \*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: \*

25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT: \*

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	369	99.5	371	24	AA170920	Sentinel Virus II
2	45	12.1	350	13	AAQ21833	Randomising oligo
3	45	12.1	350	14	AAQ36859	PCR primer for 5'
4	45	12.1	350	22	AAE76910	Sequence containin
5	45	12.1	390	24	AAE72775	Oligo #7 for clonin
6	43.2	11.6	3489	21	AAH30290	Kaposi's sarcoma-
7	43.2	11.6	3489	22	AAH82901	Kaposi's sarcoma-
8	43.2	11.6	3489	24	ABR93487	Nucleotide sequen
						Kaposi's sarcoma-

C	9	43.2	11.6	32207	20	AAV78805	KSHV LTR DNA (nucleo
C	10	43.2	11.6	137507	19	AAV19941	KSHV long unique c
C	11	43	11.6	150	20	AAV64956	Mouse histone H2B
C	12	43	11.6	125157	21	AAV74202	Nucleotide sequenc
C	13	42	11.3	2254	22	AACT8059	Human cancer assoc
C	14	42	11.3	3111	23	ABL13753	Drosophila melanog
C	15	42	11.3	5253	22	AAAS00419	Human cell regulat
C	16	42	11.3	5253	24	ABN95157	Gene #155 used tce
C	17	42	11.3	8028	23	ABL13752	Drosophila melanog
C	18	41.8	11.3	1437	23	AAAS54101	Pseudomonas aerugi
C	19	40.4	10.9	726	22	AAH03407	Human CDNA clone (
C	20	40.4	10.9	2277	19	AAV13836	Homo sapiens mamma
C	21	40.4	10.9	2277	19	AAV05372	Human telomerase p
C	22	40.4	10.9	659158	25	ABX16390	Mouse high growth
C	23	39.8	10.7	2142	22	AAH15254	Human CDNA sequenc
C	24	39.8	10.7	2198	22	AAH14344	Human CDNA sequenc
C	25	39.8	10.7	2392	22	AAH61221	Mouse DNA demethyl
C	26	39.8	10.7	77536	21	AAAI4651	Nucleotide sequenc
C	27	39.6	10.7	5387	24	AAAS39115	Human lung-specific
C	28	39.4	10.6	14705	23	AAAS59525	Protonibacterium
C	29	38.8	10.5	434	25	ABX37127	Bovine EST associa
C	30	38.8	10.5	1476	17	AAAT08578	Zinc finger protease
C	31	38.8	10.5	124884	22	AAH74201	Nucleotide sequenc
C	32	38.4	10.4	271	24	ABN26793	Human OREF polynuc
C	33	38.4	10.4	2787	25	AAAD51569	Human structural a
C	34	38.4	10.4	2946	25	ABRT21014	Aspergillus fumig
C	35	38.4	10.4	3000	25	ABRT19194	Aspergillus fumig
C	36	38.4	10.4	3169	25	ABRT18600	Aspergillus fumig
C	37	38.4	10.4	3169	25	ABRT20416	Aspergillus fumig
C	38	38.4	10.4	3290	22	AAAS00032	Human ATLAS-1-enc
C	39	38.4	10.4	5168	25	ABRT18006	Aspergillus fumig
C	40	38.4	10.4	5169	25	ABRT19820	Aspergillus fumig
C	41	38.4	10.4	48300	22	AAAF61281	N. magdali bacteri
C	42	38.2	10.3	14477	23	ABLO96313	Drosophila melanoc
C	43	38.2	10.3	4106	23	ABLO96342	Drosophila melanoc
C	44	38.2	10.3	349580	24	ABQ81846	Bifidobacterium lo
C	45	38	10.2	774	24	ABT10372	Human breast cancer

## ALIGNMENTS

## RESULT 1

ID AAI70920 standard; DNA; 371 BP

AC AAI70920;

DT 12-MAR-2002 (first entry)

Sentinel Virus II positive strand

SVII; H101.c33; hepatitis virus; infection; therapy; diagnosis;

2000

XX  
XX  
XX

FT	key	Location/Qualifiers
FT	CDS	1..370
FT		/*tag= a
FT		/partial
FT		/note= "Xaa=unknown; ORF P1, encodes AAM50524"
FT	unsure	14..16
FT		/*tag= b
FT		/note= "encodes Phe or Leu"
FT	unsure	41..43
FT		/*tag= c
FT		/note= "encodes Met or Leu"
FT	CDS	complement (1..371)
FT		/*tag= d
FT		/partial
FT		/trans_except= "(pos:96..98,aa:Xaa)"
FT		/note= "Xaa=unknown; ORF M1, encodes AAM50525"

FT	unSURE	/complement (3..5)
FT		/*tag= e
FT		/note= "encodes Gly or is termination codon"
FT	unSURE	/complement (12..14)
FT		/*tag= f
FT		/note= "encodes Ser or Asn"
FT	unSURE	/complement (15..17)
FT		/*tag= g
FT		/note= "encodes Gln or His"
FT	unSURE	/complement (39..41)
FT		/*tag= h
FT		/note= "encodes Ala or Pro"
FT	CDS	/complement (1..75)
FT		/*tag= i
FT		/partial
FT	unSURE	/note= "ORF M2, encodes AAM50526"
FT		/complement (4..6)
FT		/*tag= j
FT		/note= "encodes Leu or Arg"
FT	unSURE	/complement (50..42)
FT		/*tag= k
FT		/note= "encodes Ser or Thr"
FT	unSURE	/complement (1)
FT		/*tag= l
FT		/note= "there may can be an additional A, T, G, or C at the 3' end of the negative strand"
PX	MO200185770-A2.	
XX		
XX	15-NOV-2001.	
PD		
XX		
PF	04-MAY-2001; 2001WO-EP05029.	
PR		
XX	05-MAY-2000; 2000US-202271P.	
PA	(HOFF ) ROCHE DIAGNOSTICS GMBH. (HOFF ) HOFFMANN LA ROCHE & CO AG F.	
PA		
PI	Liu J, Lewis S, Bohenzky R, Lin Y, Ramaswamy L, Montiel J; Batz H, Chen B;	
PI		
XX		
XX	WPI: 2002-062234/08.	
DR	P-PsDB; AAM50524, AAM50525, AAM50526.	
XX		
XX	A new virus associated with cryptogenic non-A/non-G hepatitis, designated Sentinel Virus II is useful to detect and treat Sentinel Virus II infection -	
PT		
PS	Claim 2; Page 35; 20pp; English.	
XX		
CC	The present sequence is that of the positive strand of a new hepatitis virus, designated Sentinel virus II (SVII) or H101.c33, that is associated with cryptogenic, non-A/non-G hepatitis. The CC protoytic virus comprises a DNA genome of at least 371 baases. DNA CC clones comprising the genomic DNA were isolated using a modification of the representation different analysis (RDA) method in which serum from a cryptogenic hepatitis patient was used as the source of tester DNA. 3 Putative open reading frimess were identified. 1 on the positive strand (pi) of the genomic DNA and 2 CC on the negative strand (M1 and M2). Conceptual translations of these ORFs are given in AAM50524-26. SVII was found at a high CC prevalence in serum from acute hepatitis patients, and was also found in serum samples from chronic hepatitis patients, particularly hepatitis C virus patients and patients superinfected with more than one type of hepatitis virus. The invention CC provides SVII vinnuses, polynucleotides, proteins, and antibodies which bind SVII virus and viral proteins. These can be used to detect SVII virus or infection by SVII virus, and the polynucleotides can be inserted into recombinant expression vectors for production of viral proteins. Vaccines for prevention and/or treatment of SVII infection are also provided. These may be protein- or DNA-based vaccines.	

[illegible]

```

RESULT 2
AAQ21833
ID   AAQ21833 standard; DNA, 390 BP.
XX
XX   AAQ21833;
AC
XX
XX
XX
XX   08-JUN-1992 (first entry)
DT
XX
XX
XX   Randomising oligonucleotide used in SPERT mRNA prepn.
DE
XX
XX   Systematic polypeptide evolution by reverse translation; SPERT;
KW   ligand binding; ss.
XX
XX   Synthetic.
OS
XX
XX   MO9202536-A.
PN
XX
XX   20-FEB-1992.
PD
XX
XX   01-AUG-1991; 91MO-US05463.
PF
XX
XX   02-AUG-1990; 90US-0561968.
PR
XX
XX   (COLS ) UNIV OF COLORADO.
PA
XX
XX   Gold L, Tuerk C,
PI
XX
XX   WPI, 1992-080018/10.
DE
XX
XX
XX   New method of systematic polypeptide evolution by reverse
XX   translation - by linking each polypeptide in sample mixt. to
XX   individualised mRNA allowing further synthesis of selected
XX   polypeptide(s)
XX
XX   Example; Page 55; 102pp; English.
XX
XX
XX   The sequence is that of an example randomising oligonucleotide which
XX   is used in the prepn. of mRNA encoding candidate polypeptides for the

```

CC and a polypeptide encoded by its associated mRNA. The copolymers are

CC partitioned relative to their affinity to the target. The method is  
CC termed systematic polypeptide evolution by reverse translation (SPERT).  
CC The polypeptides ligands of small molecule targets are useful in assay  
CC methods, diagnostic procedures, cell sorting, as inhibitors of target  
CC molecule function, as probes, as drug delivery vehicles and modifiers of  
CC hormone action and have therapeutic uses as sequestering agents. The  
CC target molecules include natural and synthetic polypeptides, including  
CC proteins, hormones, receptors and cell surfaces, nucleic acids and small  
CC molecules such as drugs, metabolites, cofactors and toxins. Polypeptide  
CC ligands are isolated and rapidly identified by this method.  
SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;  
Query Match 12.1%; Score 45; DB 22; Length 390;  
Best Local Similarity 57.8%; Pred. No. 0.021;  
Matches 78; Conservative 1; Mismatches 56; Indels 0; Gaps 0;  
QY 21 GGTGCATGCAAGAGCGGWTGAAGCGGCGGATTTGACGACGACGACATTGCGA 80  
1 GGGCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
Db 61 TGAAGATGGAGCCGCTGACGCTCTTGCGGGGCGGAGCCGAGACCAAGACGACGAG 140  
QY 141 ACCGAGACGCTTACG 155  
121 ACCGAGACGACGACG 135  
Db 121 ACCGAGACGACGACG 135  
RESULT 5  
AA172775  
ID AA172775 standard; DNA; 390 BP.  
AC AA172775;  
DT 22-JUL-2002 (first entry)  
XX  
DE Oligo #7 for cloning T7 promoter and RBS containing mRNA.  
XX  
KW T7 promoter; ribosome binding site; RBS; prokaryotic; eukaryotic;  
KW ribosome; mRNA; circle-solid.polypeptide copolymer; mbs; PC; SPERT;  
KW Systematic Polypeptide Evolution by Reverse Translation; assay;  
KW diagnosis; cell sorting; inhibitor; probe; sequestering agent;  
KW ribosome complex; ss.  
XX  
OS Synthetic.  
XX  
PN US2002038000-A1.  
XX  
PD 28-MAR-2002.  
XX  
PF 22-FEB-2001; 2001US-0790399.  
XX  
PR 23-NOV-1998; 98US-0197649.  
PR 02-AUG-1990; 90US-0561968.  
PR 01-AUG-1991; 91US-0739055.  
XX  
PA (GOLD/) GOLD L.  
PA (TUER/) TUERK C.  
PA (PRIB/) PRIENOW D.  
PA (SMIT/) SMITH J D.  
XX  
PI Gold L, Tuerk C, Prienow D, Smith JD;  
XX  
DR WPI; 2002-329128/36.  
XX  
PT New methods (termed SPERT (Systematic Polypeptide Evolution by Reverse  
PT Translation)) for selecting high-affinity polypeptide ligands that  
PT specifically bind target molecules, e.g. proteins, carbohydrates,  
PT toxins, drugs and receptors -  
XX  
Example 1; Page 21; 38pp; English.

XX The sequences given in AA172769-81 are oligonucleotides which were used  
CC to make mRNA encoding a candidate polypeptide, a 5' fixed sequence  
CC composed of a T7 promoter sequence and a ribosome binding site which is  
CC recognised by both prokaryotic and eukaryotic ribosomes, terminating in  
CC a restriction endonuclease site. The resulting nucleic acid was used in  
CC the method of the invention for preparing ligands of target molecules  
CC in which mixtures of ribosome complexes or mRNA, circle-solid.polypeptide  
CC copolymers (mbs.pcs) are partitioned by affinity to the target and  
CC amplified to create candidate mixtures enriched in ribosome complexes  
CC or mbs.pcs with an affinity to the target, are new. The methods are  
CC termed SPERT (Systematic Polypeptide Evolution by Reverse Translation).  
CC The SPERT methods are useful for isolating polypeptide ligands for  
CC desired target molecules. The polypeptide products are useful for  
CC any purpose to which a binding reaction may be put, for example in  
CC assay methods, diagnostic procedures, cell sorting, as inhibitors of  
CC target molecule function, as probes, as sequestering agents and the  
CC like. In addition, polypeptide products of the invention can have  
CC catalytic activity. Target molecules include natural and synthetic  
CC polymers, including proteins, polysaccharides, glycoproteins, hormones,  
CC receptors and cell surfaces, nucleic acids, and small molecules such as  
CC drugs, metabolites, cofactors, transition state analogues and toxins.  
CC The novel SPERT method involves utilizing a candidate mixture of  
CC polypeptides having a randomized amino acid sequence. Each member of  
CC the mixture is linked to an individualized mRNA, which encodes the  
CC amino acid sequence of that polypeptide. The candidate polypeptides are  
CC partitioned according to their property of binding to a given desired  
CC target molecule. The partitioning is carried out in such a way that  
CC each mRNA encoding a polypeptide is partitioned exactly together with  
CC that polypeptide. In this way each polypeptide is partitioned together  
CC with the means for further amplifying it by an in vitro process.  
CC Ultimately, both the desired optimal polypeptide ligand of the desired  
CC target and the mRNA encoding the polypeptide are simultaneously  
CC selected, allowing further synthesis of the selected polypeptide as  
CC desired, and further amplification of the coding sequence. It is  
CC therefore not necessary to analyse the amino acid sequence of the  
CC selected polypeptide (using protein chemistry) in order to produce it  
CC in desired quantities.  
SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;  
Query Match 12.1%; Score 45; DB 24; Length 390;  
Best Local Similarity 57.8%; Pred. No. 0.021;  
Matches 78; Conservative 1; Mismatches 56; Indels 0; Gaps 0;  
QY 21 GGTGCATGCAAGAGCGGWTGAAGCGGCGGATTTGACGACGACGACATTGCGA 80  
1 GGGCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
Db 81 TGAAGATGGAGCCGCTGACGCTCTTGCGGGGCGGAGCCGAGACCAAGACGACGAG 140  
61 ACCGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 120  
QY 141 ACCGAGACGCTTACG 155  
121 ACCGAGACGACGACG 135  
Db 121 ACCGAGACGACGACG 135  
RESULT 6  
AAA30290  
ID AAA30290 standard; DNA; 3489 BP.  
XX  
AC AAA30290;  
XX  
DT 11-SEP-2000 (first entry)  
XX  
DE Kaposi's sarcoma-associated herpesvirus LANA gene.  
XX  
KW Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;  
KW latency-associated nuclear antigen; LANA; gamma-2 herpes virus;  
KW Human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;  
KW Kaposi's sarcoma; primary effusion lymphoma; PEL;  
KW human immunodeficiency virus; HIV; multicentric Castleman's disease; ds.

```

XX OS Kaposi's sarcoma-associated herpesvirus.
XX FH Key
XX CDS 1..3489
XX FT /tag= a
XX FT /product= "LANA"
XX FT misc_signal
XX FT /note= b
XX FT /note= "nuclear localisation signal, NLS"
XX FT misc_signal
XX FT /tag= c
XX FT /note= "nuclear localisation signal, NLS"
XX PD WO200029626-A1.
XX PD 25-MAY-2000.
XX PF 19-NOV-1999; 99WO-US27508.
XX PR 19-NOV-1998; 98US-0109422.
XX PR 21-APR-1999; 99US-0298568.
XX PA (KIEF/) KIEFF E D.
XX PA (BALU/) BALLESTAS M E.
XX PA (KAYE/) KAYE K M.
XX PI Kieff ED, Ballestas ME, Kaye KM;
XX WPI: 2000-387829/33.
XX DR P-PSDB; AAY96255.
XX PT Treating or preventing a disease associated with rhodino virus
XX PT infection in a mammal which includes Kaposi's Sarcoma and Primary
XX PT Effusion Lymphoma
XX PS Disclosure; Fig 6; 70pp; English.
XX CC The present sequence is the Kaposi's sarcoma-associated herpesvirus,
XX CC (KSHV) latency-associated nuclear antigen (LANA) gene. KSHV is also known
XX CC as Human Herpes virus 8 (HHV8) and belongs to the rhadino virus, or
XX CC gamma-2 herpes virus class. The LANA protein is necessary for the
XX CC efficient persistence of rhadino virus DNA in mammalian cells. Persistent
XX CC rhadino virus infection is implicated in a variety of diseases e.g.
XX CC Kaposi's Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
XX CC Caetlemann's disease. In addition, KS is a common malignancy in HIV
XX CC patients. KSHV persists in host cells in a latent form. One of the few
XX CC genes expressed from the latent viral DNA is LANA. LANA associates with
XX CC both human chromosomes and with the rhadino virus cis-acting element
XX CC (RVAB), thereby providing a tethering function: the KSHV DNA episome is
XX CC "tied" to the host chromosomes. This allows the viral DNA to persist in
XX CC the host cell. The present sequence may be used to screen and identify
XX CC molecules that inhibit LANA interaction with RVAB, thereby interfering
XX CC with the latency cycle of this virus. Potential antiviral treatments for
XX CC the above mentioned diseases may therefore be based on LANA deregulation.
XX SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;

Query Match 11.6%; Score 43.2; DB 21; Length 3489;
Best Local Similarity 59.0%; Pred. No. 0.11;
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;
QY 28 GCAGAGACCGGWTGAAGCGGACGATTTGACGACGACGACGACGATTCGATGAAGA 87
DB 1086 GGAGATATGACGAGGATGACGAGAGATGACGAGATGACGAGAGATGACGA 1145
QY 88 TGGGACCGCGYACGTCCTTGGCGGGCGGAGCGGAGAACCAAGACGACGAGAGAGA 147
DB 1146 GGAGATATGACGAGGAGAGACGAGAGAGACGAGACGAGACGAGAGAGAGA 1205
QY 148 CG 149
DB 1206 CG 1207

```

```

RESULT 7
AAF82901
ID AAF82901 standard; DNA; 3489 BP.
XX AC AAF82901;
XX DT 29-JUN-2001 (first entry)
XX DE Nucleotide sequence of KSHV tethering protein, LANA.
XX KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
XX KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
XX KW KSHV; latency-associated nuclear antigen; LANA; ds.
XX OS Kaposi's sarcoma associated herpesvirus.
XX FH Key
XX CDS 1..3489
XX FT /tag= a
XX FT WO200125484-A2.
XX PD 12-APR-2001.
XX PF 29-SEP-2000; 2000WO-US26908.
XX PR 01-OCT-1999; 99US-0410399.
XX PA (UNMI ) UNIV MICHIGAN.
XX PI Robertson ES, Cotter MA;
XX WPI: 2001-281736/29.
XX DR P-PSDB; AAB62331.
XX PT A composition for use in gene therapy comprises an expression vector
XX PT that includes a nucleic acid sequence encoding a nucleic acid binding
XX PT protein
XX PS Disclosure; Fig 9A; 60pp; English.
XX CC The invention provides a composition comprising nucleic acid, histone H1
XX CC protein and expression vector operationally encoding a protein suitable
XX CC for tethering the nucleic acid to the histone H1 protein, where the
XX CC tethering protein is LANA. The composition is useful in aiding the
XX CC retention of the viral DNA in the host cell. The viral vector encodes a
XX CC protein suitable for tethering DNA to Histone H1. Methods for screening
XX CC for compounds which are agonistic or antagonistic for the tethering of
XX CC viral proteins to histone H1 and DNA binding sites are useful for
XX CC developing the method of viral transfer. The composition has applications
XX CC to gene therapy, including the treatment of multiple sclerosis,
XX CC Parkinson's disease, Huntington disease and diabetes. The present
XX CC sequence represents the nucleotide sequence of the Kaposi's sarcoma
XX CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
XX CC antigen (LANA), which acts as a tethering protein.
XX SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;

Query Match 11.6%; Score 43.2; DB 22; Length 3489;
Best Local Similarity 59.0%; Pred. No. 0.11;
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;
QY 28 GCAGAGACCGGWTGAAGCGGACGATTTGACGACGACGACGACGATTCGATGAAGA 87
DB 1086 GGAGATATGACGAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGA 1145
QY 88 TGGGACCGCGYACGTCCTTGGCGGGCGGAGCGGAGAACCAAGACGACGAGAGAGA 147
DB 1146 GGAGATATGACGAGGAGAGACGAGAGAGACGAGAGAGACGAGAGAGAGAGA 1205
QY 148 CG 149

```







XX AAV64956;  
 AC  
 XX  
 DT 05-FEB-1999 (first entry)  
 XX  
 DE Mouse histone H2B protein epitope oligonucleotide fragment #7.  
 XX  
 KW ligand; assay; diagnostic; inhibitor; probe; cell sorting; catalysis;  
 KM sequestering agent; epitope; ss.  
 XX  
 OS Synthetic.  
 XX Mus sp.  
 PN US5843701-A.  
 XX  
 PD 01-DEC-1998.  
 XX  
 PF 31-JAN-1992; 92US-0829461.  
 XX  
 PR 31-JAN-1992; 92US-0829461.  
 PR 02-AUG-1990; 90US-0561968.  
 PR 01-AUG-1991; 91US-0739055.  
 XX  
 PA (NEXS-) NEXSTAR PHARM INC.  
 XX  
 PI Gold L, Pribnow D, Smith JD, Tuerk C;  
 DR WPI; 1999-044575/04.  
 XX  
 PT Production of polypeptide ligands using mRNA-polypeptide conjugates  
 PT - by affinity selection, RNA amplification and polypeptide synthesis  
 XX  
 PS Example 10; Column 45-46; 35bp; English.  
 XX  
 CC AAV64950-V64972 are sequences used in a method for the production of a  
 CC polypeptide ligand of a target molecule. The polypeptide products can  
 CC be used in assays, in diagnostic procedures, in cell sorting, as  
 CC inhibitors of target molecule function, as probes, as sequestering  
 CC agents, or may have catalytic activity.  
 CC  
 XX Sequence 150 BP; 44 A; 46 C; 53 G; 6 T; 1 other;  
 SQ  
 Query Match 11.6%; Score 43; DB 20; Length 150;  
 Best Local Similarity 58.4%; Pred. No. 0.057;  
 Matches 73; Conservative 1; Mismatches 51; Indels 0; Gaps 0;  
 QY 32 AAGACGGGWTGAAGCGGACGGATTGACGACGACGACATTGCGATGAAGATGG 91  
 DB 6 ATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 65  
 QY 92 ACCGCTGACGCTCTTTGGGGGGCGGAGCGGAGAACGACGACGACGACGACGCTC 151  
 DB 66 GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 125  
 QY 152 TACGC 156  
 DB 126 GACGC 130  
 DE  
 XX Nucleotide sequence of the Varicella virus Oka strain.  
 XX  
 KM Oka strain; pox vaccine; vaccine; poxvirus disease; chickenpox; ss.  
 XX  
 OS Varicella virus.  
 XX

PN WO200156600-A1.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 31-JAN-2001; 2001WO-IP00678.  
 XX  
 PR 31-JAN-2000; 2000JP-0062734.  
 XX  
 PA (OSAU ) UNIV OSAKA.  
 XX  
 PI Gomi Y, Sunamachi H, Takahashi M, Yamanihi K;  
 DR WPI; 2001-488845/53.  
 XX  
 PT Method for checking absence of mutation at specific positions of  
 PT varicella virus genome for quality control of attenuated live varicella  
 PT vaccine -  
 XX  
 PS Disclosure; Page 106-150; 158bp; Japanese.  
 XX  
 CC The present sequence represents the nucleotide sequence of Varicella  
 CC virus, Oka strain. The specification describes a method for the  
 CC quality control of an attenuated pox vaccine, characterised in that  
 CC the DNA sequence of specific parts of the viral genome in a sample is  
 CC determined and proved to be conserved rather than mutated in comparison  
 CC with the same parts of a reference viral genome. The method is useful  
 CC for quality control of vaccines for use in the prevention of poxvirus  
 CC diseases such as chickenpox.  
 CC  
 XX Sequence 125157 BP; 33776 A; 29439 C; 28285 G; 33624 T; 33 other;  
 SQ  
 Query Match 11.6%; Score 43; DB 22; Length 125157;  
 Best Local Similarity 58.4%; Pred. No. 0.27;  
 Matches 73; Conservative 1; Mismatches 51; Indels 0; Gaps 0;  
 QY 33 AGACCGGWTGAAGCGGACGGATTGACGACGACGACGACATTGCGATGAAGATGG 92  
 DB 14043 ACCACCGAGGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 14102  
 QY 93 CCGCTGACGCTCTTTGGGGGGCGGAGCGGAGAACGACGACGACGACGACGCTC 152  
 DB 14103 TCGACGACGAGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 14162  
 QY 153 ACGCG 157  
 DB 14163 ACGCG 14167  
 DE  
 XX  
 RESULT 13  
 AAC78059  
 ID AAC78059 standard; cDNA; 2254 BP.  
 XX  
 AC AAC78059;  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE  
 XX Human cancer associated gene sequence SEQ ID NO:453.  
 XX  
 KM Human; cancer associated gene; cancer antigen; detection; cancer;  
 KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 KM antidiabetic; antisthmatic; antineumatic; antiarthritic; antiviral;  
 KM antinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;  
 KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
 KM vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 KM immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KM allergic reaction; graft versus host disease; organ rejection;  
 KM haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KM neurological disease; drug screening; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055350-A1.  
 XX

PD 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US05882.  
 PF 12-MAR-1999; 99US-0124270.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Ruben SM;  
 XX WPI: 2000-587533/55.  
 DR P-PSDB; AAB43850.  
 XX Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer -  
 XX  
 PS Claim 1; Page 990-991; 2352pp; English.  
 XX AACT77607 to AACT8448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antiasthmatic; antineumatic; antiarthritis;  
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;  
 CC neurotropic; vasotropic; antipsoriatic and angiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AACT8449 to  
 CC AACT8457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 2254 BP; 491 A; 657 C; 654 G; 447 T; 5 other;  
 S0  
 Query Match 11.3%; Score 42; DB 21; Length 2254;  
 Best Local Similarity 59.0%; Pred. No. 0.2;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 33 AGGAGGGGTGAAGCGGACGGGATTGACGACGACGACGACATTGGGATGAAGAATGGGA 92  
 DB 1088 AGGTGGGCTCTTGGCGGCGCACTGGCAAGGACACCAAGAGCGAAGGAGGAAAGCTGGGG 1147  
 QY 93 CCGGTCGAGCTCCTTGGCGGGGCGGAGCGGAGAACCAAGACGACGAGCGAGGAGCTCT 152  
 DB 1148 CAGCGACGACGTGTGATGTGGCGAGGAGAGGAGGAGGAGGAGGATGTAGGACTCT 1207  
 QY 153 AC 154  
 DB 1208 AC 1209  
 RESULT 14  
 ABL13753  
 ID ABL13753 strand; cDNA; 3111 BP.  
 XX ABL13753;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35741.  
 DE  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmacological; gene; ss.  
 XX Drosophila melanogaster.  
 OS

XX WO200171042-A2.  
 PN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI: 2001-656860/75.  
 DR P-PSDB; ABB69650.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 35741; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (AB101840-ABL16175) and the encoded proteins  
 CC (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pt\_sequences.  
 XX  
 SQ Sequence 3111 BP; 944 A; 791 C; 876 G; 500 T; 0 other;  
 S0  
 Query Match 11.3%; Score 42; DB 23; Length 3111;  
 Best Local Similarity 57.1%; Pred. No. 0.21;  
 Matches 72; Conservative 2; Mismatches 52; Indels 0; Gaps 0;  
 QY 30 AGAAGGACCGGTGAAGCGGCGGACGGGATTGACGACGACGACGACATTGGGATGAAGAATG 89  
 DB 2290 AAGAGAGAAACAAGACGAGGAGGCGGATGAGAGACGATCAGATGACAAAAAGGCCG 2349  
 QY 90 GGACCGCTGACGCTCCTTGGCGGGGCGGAGCGGAGAACCAAGACGACGAGCGAGAGC 149  
 DB 2350 GTGATGATGACCGCCGATGCGCATGAGACATTGATGCAAGTTAAACGAGATGACGACG 2409  
 QY 150 TCTACG 155  
 DB 2410 AGTACG 2415  
 RESULT 15  
 AAS00419  
 ID AAS00419 strand; cDNA; 5253 BP.  
 XX AAS00419;  
 AC  
 XX 09-MAY-2001 (first entry)  
 DT  
 XX Human cell regulatory protein p193 cDNA.  
 DE  
 XX Human; cell regulatory protein; p193; apoptosis associated protein;  
 KW BH3 domain; cell cycle; cardiomyocyte; tumorigenic cell;  
 KW apoptotic activity; cell proliferation; anti-apoptotic;  
 KW myocardial infarct; cardiomyopathy; cardiovascular disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 87..5183

```

FT      /*tag= a
FT      /product= "p193 protein"
XX
XX      W0200114418-A2.
XX
XX      01-MAR-2001.
XX
XX      23-AUG-2000; 2000MO-US23161.
XX
XX      23-AUG-1999; 99US-0150266.
XX
XX      (ADRE-) ADVANCED RES & TECHNOLOGY INST.
XX
XX      Field LJ, Tsai S;
XX
XX      WPI; 2001-218429/22.
XX
XX      P-PsDB; AAU00416.
XX
XX      New apoptosis-associated p193 protein, used as a target for modifying
XX      the cell cycle involving modifying the p193 protein level and/or
XX      interfering with the p193 signal transduction pathway to induce or
XX      suppress apoptosis
XX
XX      Claim 26; Page 83-91; 95pp; English.
XX
XX      The present sequence encoding for novel human cell regulatory protein
XX      p193 is isolated from an adult heart cDNA library. The p193 protein
XX      is an apoptosis-associated protein comprising a BH3 domain (AAU00417).
XX      A homologous mouse p193 protein (AAU00415) is also described. The p193
XX      protein is useful as a target for modifying the G1 phase of the cell
XX      cycle of a mammalian cell, preferably human and for inducing apoptosis
XX      in an inappropriately proliferative cell. A nucleic acid encoding p193 is
XX      useful for providing vectors for functionally introducing the p193
XX      polynucleotide sequence into mammalian or other eukaryotic cells such as
XX      cardiomyocytes, hepatocytes, smooth muscle cells, haematopoietic stem
XX      cells or tumorigenic cells. A host cell comprising p193 polynucleotide
XX      is useful in the study and understanding of the cell cycle, in screening
XX      for chemical or physical agents which modulate p193 activity or other
XX      aspects of the cell cycle, or in the culture of cells having suppressed
XX      apoptotic activity and/or increased proliferative potential for
XX      subsequent engraftment into mammals, including humans. The cells
XX      preferably left ventricular, right ventricular, right and left arterial
XX      cardiomyocytes or a mixture of all these cells are genetically modified
XX      in vitro to incorporate anti-apoptotic and/or proliferative p193 nucleic
XX      acids which are delivered to the subjects in a damaged or diseased area
XX      of the heart in cases of myocardial infarcts and cardiomyopathies.
XX
XX      Sequence 5253 BP; 1109 A; 1531 C; 1564 G; 1049 T; 0 other;
XX
XX      Query Match      11.3%; Score 42; DB 22; Length 5253;
XX      Best Local Similarity 59.0%; Pred. No. 0.24;
XX      Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
XX
XX      QY      33 AGACGGGWTGAAGCGGACGATTGACGACGACGACATTGCGATGAAGATGGGA 92
XX      DB      4108 AGGTGGGCTTGGGGCCAGTGCACAGAGCACAAGAGGAGAGAGAGAGCTGGGG 4167
XX
XX      QY      93 CCGCYGACGTCCTTGGCGGGCGGAGCGCGGAAACCAAGACGACGAGACGACGCTCT 152
XX      DB      4168 CAGCAGCAGTGTGTGATGTGGCGGAGGAGAGAGAGAGAGAGATGAGACCTCT 4227
XX
XX      QY      153 AC 154
XX      DB      4228 AC 4229

```

Search completed: September 14, 2003, 23:24:02  
 Job time : 224 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 23:16:14 ; Search time 55 Seconds  
(without alignments)  
2977.330 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371  
Sequence: 1 gatcmgaacgcytsctc.....ctctctaccgcagcagc 371

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/pdata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/pdata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/pdata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/pdata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/pdata/2/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/pdata/2/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	50.2	13.5	1926	4	US-09-249-585A-4
C 2	50.2	13.5	1931	2	US-09-130-114-2
C 3	45	12.1	390	3	US-09-197-649-7
C 4	43.4	11.7	789	4	US-09-252-991A-9465
C 5	43.4	11.7	918	4	US-09-252-991A-9337
C 6	43.4	11.7	1440	4	US-09-252-991A-9411
C 7	43.2	11.6	3489	2	US-08-728-322A-1
C 8	43.2	11.6	3489	4	US-09-298-568-1
C 9	43.2	11.6	3489	4	US-09-410-399-1
C 10	43.2	11.6	32207	2	US-08-770-379-20
C 11	43.2	11.6	32207	4	US-08-757-669A-20
C 12	43.2	11.6	32207	4	US-09-230-371A-20
C 13	43	11.6	150	2	US-07-829-461A-8
C 14	41.2	11.1	1146	4	US-09-252-991A-10603
C 15	41.2	11.1	1260	4	US-09-252-991A-10250
C 16	40.8	11.0	7218	1	US-08-232-463-14
C 17	40.4	10.9	2277	1	US-08-676-967-5
C 18	40.4	10.9	2277	1	US-08-676-974-5
C 19	40.4	10.9	2277	1	US-09-098-487-5
C 20	39.8	10.7	77536	4	US-09-110-551B-1
C 21	39	10.5	468	4	US-09-252-991A-1203
C 22	39	10.5	1275	4	US-09-252-991A-1311
C 23	39	10.5	3981	4	US-09-252-991A-1361
C 24	39	10.5	4179	4	US-09-252-991A-1244
C 25	38.8	10.5	124884	4	US-09-661-596A-76
C 26	38.6	10.4	1119	4	US-09-252-991A-10788
C 27	38.2	10.3	1041	4	US-09-252-991A-14894

C 28	38.2	10.3	1170	4	US-09-252-991A-15019	Sequence 15019, A
C 29	38.2	10.3	1692	4	US-09-252-991A-14477	Sequence 14477, A
C 30	38	10.2	1865	3	US-09-328-869-1	Sequence 1, Appli
C 31	38	10.2	1865	3	US-09-629-774A-1	Sequence 1, Appli
C 32	37.6	10.1	960	1	US-07-639-330E-1	Sequence 1, Appli
C 33	37.6	10.1	1095	1	US-07-639-330E-2	Sequence 2, Appli
C 34	37.2	10.0	2817	4	US-09-620-312D-1085	Sequence 1085, Ap
C 35	37	10.0	276	4	US-09-252-991A-12547	Sequence 12547, A
C 36	36.4	9.8	1464	4	US-08-528-199-2	Sequence 2, Appli
C 37	36.4	9.8	1704	1	US-08-528-199-5	Sequence 12936, A
C 38	36.4	9.8	1854	4	US-09-252-991A-12936	Sequence 1150, Ap
C 39	36.2	9.8	444	4	US-09-252-991A-1150	Sequence 1239, Ap
C 40	36.2	9.8	1410	4	US-09-252-991A-1239	Sequence 1316, Ap
C 41	36.2	9.8	1431	4	US-09-252-991A-1316	Sequence 2, Appli
C 42	36.2	9.8	2277	1	US-08-676-967-2	Sequence 2, Appli
C 43	36.2	9.8	2277	1	US-08-676-974-2	Sequence 2, Appli
C 44	36.2	9.8	2277	2	US-09-098-487-2	Sequence 2, Appli
C 45	36.2	9.8	2277	2	US-09-098-487-2	Sequence 2, Appli

## ALIGNMENTS

```
RESULT 1
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/00905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match      13.5%  Score 50.2;  DB 4;  Length 1926;
Best Local Similarity 62.2%;  Pred. No. 0.00013;
Matches 79;  Conservative 0;  Mismatches 48;  Indels 0;  Gaps 0;

QY      31 GAAGACGGGWTGAAGCGGACGCGGATTGACGACGACGACATTTGGATGAAGATGG 90
      |||||
DB      929 GGAGACCGGGAGGACGAGACGCGGAGAGACCGGAGACGAGACGCGGAGACGCGGGA 870
      |||||

QY      91 GACCGCYAGCTCTTGGCGGGCGGAGCGGAGACCAACGACGACGACGACGCT 150
      |||||
DB      869 GGACGACGACGCGGAGGACGCGGAGGAGACGAGACGCGGAGGAGACGAGAGACGA 810
      |||||

QY      151 CTACGCG 157
      |||||
DB      809 GGACGCGG 803

RESULT 2
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damej, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/ID903US1
```

;; CURRENT APPLICATION NUMBER: US/09/130.114  
;; CURRENT FILING DATE: 1998-08-06  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 1931  
;; TYPE: DNA  
;; ORGANISM: EBNA  
US-09-130-114-2

Query Match 13.5%; Score 50.2; DB 2; Length 1931;  
Best Local Similarity 62.2%; Pred. No. 0.00013;  
Matches 79; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 31 GAAGACGGGWTGAAGCGGATTCAGACGACGACGACGATTCGATGAAGATGG 90  
DB 929 GAGAGACGGGAGAGACGAGGACGCGGAGAGACGAGGAGACGCGGAGCGGGA 870  
QY 91 GACCGCTGACGCTCTTGGCGGCGGCGGAGCGGAGAACCAAGACGAGGACGAGCGT 150  
DB 869 GAGACGAGGACGCGGAGAGACGCGGAGAGACGAGACGAGGAGGAGCGGAGGAGCGA 810  
QY 151 CTACGCG 157  
DB 809 GAGCGG 803

## RESULT 3

US-09-197-649-7  
Sequence 7, Application US/09197649

;; Patent No. 6194550  
;; GENERAL INFORMATION:  
;; APPLICANT: Gold, Larry  
;; APPLICANT: Tuerk, Craig  
;; APPLICANT: Pridnow, David  
;; APPLICANT: Smith, Jonathan D.  
;; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation  
;; FILE REFERENCE: NEX02/C1-CON  
;; CURRENT APPLICATION NUMBER: US/09/197,649  
;; EARLIER FILING DATE: 1998-11-23  
;; EARLIER APPLICATION NUMBER: 07/829,461  
;; EARLIER FILING DATE: 1992-01-31  
;; EARLIER APPLICATION NUMBER: 07/739,055  
;; EARLIER FILING DATE: 1991-08-01  
;; EARLIER APPLICATION NUMBER: 07/561,968  
;; EARLIER FILING DATE: 1990-08-02  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: Patent In Ver. 2.0  
;; SEQ ID NO 7  
;; LENGTH: 390  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Sequence  
;; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed  
;; OTHER INFORMATION: fragments having NcoI restriction sites.  
US-09-197-649-7

Query Match 12.1%; Score 45; DB 3; Length 390;  
Best Local Similarity 57.8%; Pred. No. 0.002;  
Matches 78; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 21 GGTGATGAGAAAGACGGGWTGAAGCGGATTCAGACGACGACGACGATTCGCA 80  
DB 1 GGGCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
QY 81 TGAAGATGAGACCGCTGACGCTCTTGGCGGCGGAGCGGAGAACCAAGACGAGG 140  
DB 61 ACAGACGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAG 120  
QY 141 ACAGAGACGCTTACG 155  
DB 121 ACAGACGACGACGACG 135

RESULT 4  
US-09-252-991A-9465/C  
Sequence 9465, Application US/09252991A

;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 9465  
;; LENGTH: 789  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9465

Query Match 11.7%; Score 43.4; DB 4; Length 789;  
Best Local Similarity 53.5%; Pred. No. 0.0069;  
Matches 83; Conservative 3; Mismatches 69; Indels 0; Gaps 0;

QY 12 GYTTSGCTCGGTGATGACGAGAGCGGWTGAAGCGGAGCGGATTCAGACGACGAG 71  
DB 318 GCGCGCTCGGCGCGGATGACGAGCGGAGCGGAGCGGATTCGATGCGGATG 259  
QY 72 ACATTGCGATGAAGAAGTACCGGCTGACGCTTGGCGGCGGAGCGGAGAACCAAG 131  
DB 258 ACCTTACGACGCGCGGCTGCTTGGAGCATGCGGAGCGGAGCGGAGCGGAGCGG 199  
QY 132 ACAGACGAGACGAGACGCTTACGCGCGCATCCGT 166  
DB 198 AAGTGCGGACGTTGATTCGACGCGCTCGGCTT 164

## RESULT 5

US-09-252-991A-9337  
Sequence 9337, Application US/09252991A

;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 9337  
;; LENGTH: 918  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9337

Query Match 11.7%; Score 43.4; DB 4; Length 918;  
Best Local Similarity 53.5%; Pred. No. 0.0072;  
Matches 83; Conservative 3; Mismatches 69; Indels 0; Gaps 0;

QY 12 GYTTSGCTCGGTGATGACGAGAGCGGWTGAAGCGGAGCGGATTCAGACGACGAG 71  
DB 182 GCGCGCTCGGCGCGGATGACGCTTGGAGCGGAGCGGAGCGGATTCGATGCGGATG 241  
QY 72 ACATTGCGATGAAGAAGTACCGGCTTGGCGGCGGAGCGGAGAACCAAG 131

Db 242 ACCTTGACCAAGCGCGGTGCTGCTTGGACGCCATCGCGGACCGCTGGCGGCAACGGG 301  
Qy 132 ACACGAGGAGGACGAGCGTCTACCGCGGATCCGT 166  
Db 302 AAGGTCCGACGTTGACTTGAAGCGCCCTGGCCTT 336

RESULT 6  
US-09-252-991A-9411/C  
; Sequence 9411, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9411  
; LENGTH: 1440  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9411

Query Match 11.7%; Score 43.4; DB 4; Length 1440;  
Best Local Similarity 53.5%; Pred. No. 0.0083;  
Matches 83; Conservative 3; Mismatches 69; Indels 0; Gaps 0;

Qy 12 GTTSGCTGCGTGCATGCGAAGAGGACGGGWTGAAGCGGACGCGGATGACGACGAGG 71  
Db 1277 GCGGCGCTGCGGCGCATGCGTGAGCGCCAGACCGCGTGGTCTTTCGATGCGGATG 1218

Qy 72 ACATTGCGATGAAGAATGGAGCCGCGYACGTCCTTGGCGGCGGACCGCGAGAACCAAG 131  
Db 1217 ACCTTGACCAAGCGCGGTGCTGCTTGGACGCAATCGCCGACCGCTGGCGGGAACGGG 1158

Qy 132 ACACGAGGAGGACGAGCGTCTACCGCGGATCCGT 166  
Db 1157 AAGGTCCGACGTTGACTTGAAGCGCCCTGGCCTT 1123

RESULT 7  
US-08-728-323A-1  
; Sequence 1, Application US/08728323A  
; Patent No. 5948676  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus' DNA  
; TITLE OF INVENTION: Encoding Same And Uses Thereof  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,323A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 0575/52268/JBW/MSK/SKS  
; TELEPHONE: 212-278-0400  
; TELEFAX: 212-391-0525  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3489 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3489  
US-08-728-323A-1

Query Match 11.6%; Score 43.2; DB 2; Length 3489;  
Best Local Similarity 59.0%; Pred. No. 0.012;  
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

Qy 28 GCAGAGGACGGGWTGAAGCGCGGATTTGACGACGACGACGATTCGATGAAGA 87  
Db 1086 GGAAGATGACGAGGAGATGACGAGGAGATGACGAGGAGATGACGAGGAGATGACGA 1145

Qy 88 TGGACCGCGYACGTCCTTGGCGGCGGACCGCGAACCAGACGACGAGACGAGGA 147  
Db 1146 GGAGATGACGAGGAGGAGGACGAGGAGGAGGACCGAGGAGGAGGAGGAGGAGGAGGA 1205

Qy 148 CG 149  
Db 1206 CG 1207

RESULT 8  
US-09-298-568-1  
; Sequence 1, Application US/09298568  
; Patent No. 6322792  
; GENERAL INFORMATION:  
; APPLICANT: Kieff, Elliott D.  
; APPLICANT: Ballesta, Mary E.  
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE  
; FILE REFERENCE: 16412-10001R  
; CURRENT APPLICATION NUMBER: US/09/298,568  
; CURRENT FILING DATE: 1999-04-21  
; EARLIER APPLICATION NUMBER: US 60/109,422  
; EARLIER FILING DATE: 1998-11-19  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3489  
; TYPE: DNA  
; ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-298-568-1

Query Match 11.6%; Score 43.2; DB 4; Length 3489;  
Best Local Similarity 59.0%; Pred. No. 0.012;  
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

Qy 28 GCAGAGGACGGGWTGAAGCGCGGATTTGACGACGACGACGATTCGATGAAGA 87  
Db 1086 GGAAGATGACGAGGAGATGACGAGGAGATGACGAGGAGATGACGAGGAGATGACGA 1145

Qy 88 TGGACCGCGYACGTCCTTGGCGGCGGACCGCGAACCAGACGACGAGACGAGGA 147  
Db 1146 GGAGATGACGAGGAGGAGGACGAGGAGGAGGACCGAGGAGGAGGAGGAGGAGGAGGA 1205

QY 148 CG 149  
Db 1206 CG 1207

## RESULT 9

US-09-410-399-1  
Sequence 1, Application US/09410399  
Patent No. 6482587  
GENERAL INFORMATION:  
APPLICANT: Robertson, Btle S.  
APPLICANT: Cotter, Murray A.  
TITLE OF INVENTION: Methods to inhibit or enhance the binding of viral DNA  
FILE REFERENCE: UM-03778  
CURRENT APPLICATION NUMBER: US/09/410,399  
CURRENT FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 3489  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-410-399-1

Query Match 11.6%; Score 43.2; DB 4; Length 3489;  
Best Local Similarity 59.0%; Pred. No. 0.012;  
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

QY 28 GCAGAGACGCGGTGAAGCGGCGGATTCAGACGACGACATTCGCGATGAAGA 87  
Db 1086 GGAGAGATGACGAGAGCGATGACGAGAGATGACGAGAGATGACGAGAGATGACGA 1145  
QY 88 TGGAGCCGCGYACGTCCTTGGCGGGCGGACCGCGAAGACCAAGACGACGAGACGA 147  
Db 1146 GGAGAGATGACGAG 1205  
QY 148 CG 149  
Db 1206 CG 1207

## RESULT 10

US-08-770-379-20/c  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770,379  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 11.6%; Score 43.2; DB 2; Length 32207;  
Best Local Similarity 59.0%; Pred. No. 0.024;  
Matches 72; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

QY 28 GCAGAGACGCGGTGAAGCGGCGGATTCAGACGACGACATTCGCGATGAAGA 87  
Db 20911 GGAGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGA 20852  
QY 88 TGGAGCCGCGYACGTCCTTGGCGGGCGGACCGCGAAGACCAAGACGAGAGACGA 147  
Db 20851 GGAGATGACGAG 20792  
QY 148 CG 149  
Db 20791 CG 20790

## RESULT 11

US-08-757-669A-20/c  
Sequence 20, Application US/08757669A  
Patent No. 6183751  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,669A  
FILING DATE:  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double





ORGANISM: *Pseudomonas aeruginosa*  
US-09-252-991A-10603

Query Match	11.1%	Score 41.2	DB 4	length 1146
Best Local Similarity	53.4%	Pred. No. 0.031		
Matches	79	Conservative	3	Mismatches 66; Indels 0; Gaps 0;

16 SGCTGGTGCATGCAGAGGACGGGWTGAAGCCGGA

Db 36 GGTTCTGGGGGTTGAGCGTGCGCCAGCGAGGCGGACCTGAAGAGGCCCTACCGCGCCT 95

76 TGGGATGAAGATGGGACCGCYGACCTCTTGGCGGGGGGAGCGCGAGAACCAAGACGA 135

Db 96 GCGGATGAATATCACCCGGACCGCAATCCTGGCGACAAGGAGCCGAGGACAAGTTCAA 155

QY 136 CGAGGACGAGGACGTCTACGCGGCATC 163

Db 156 GGAGGCCAACGAGGCTACGAGGTCTC 183

**RESULT 15**

US-09-252-991A-10250/c  
; Sequence 10250, Application US/09252991A

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 10230
;
; LENGTH: 1260

```

```

;
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

US-09-252-991A-10250

Query Match	11.18;	Score 41.
Best Local Similarity	53.48;	Pred. No.

Matches	79;	Conservative	3;	Mismatch
---------	-----	--------------	----	----------

16 SGCTGGTGCATGCAGAAAGGACGGGTGA

Db . 1105 GGTCTGGGGGTTGAGCGTGGCGCCACGCC

985

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

Search completed: September 15, 2003, 00:34:50

0  
0  
0  
  
C  
Y  
E  
S  
  
.  
  
U  
O  
  
0  
0  
0  
0

GenCore version 5.1.6  
Copyright (c) 1993-2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2003, 00:00:35 ; Search time 172 Seconds

(without alignments)  
5236.614 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371

Sequence: 1 gatcmggaacgcttcgctc.....ctctctaccgcagcagc

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1632420 seqs, 1213878141 residues 3264840

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- Published Applications NA.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369	99.5	371	US-09-849-729-1	Sequence 1, Appli
2	366.6	98.8	372	US-09-849-729-2	Sequence 2, Appli
3	45	12.1	390	US-09-790-399-7	Sequence 7, Appli
4	43.2	11.6	3489	US-10-294-804-1	Sequence 1, Appli
5	43	11.6	261	US-09-913-514-23	Sequence 23, Appli
6	43	11.6	125157	US-09-913-514-2	Sequence 2, Appli
7	42	11.3	470	US-09-918-985-28591	Sequence 28591, A
8	42	11.3	2254	US-09-925-301-453	Sequence 453, App
9	42	11.3	5253	US-09-880-107-1655	Sequence 1655, Ap
10	42	11.3	5253	US-10-080-943-3	Sequence 3, Appli
11	41.8	11.3	1437	US-09-815-242-7738	Sequence 7738, Ap
12	41.6	11.2	498	US-10-156-761-2598	Sequence 1, Appli
13	41.6	11.2	9025608	US-10-156-761-1	Sequence 1, Appli
14	40.4	10.9	228	US-09-913-514-31	Sequence 31, Appli
15	40.4	10.9	659158	US-09-771-208-20	Sequence 20, Appli
16	39.8	10.7	494	US-09-918-995-32147	Sequence 32147, A

C	17	39.6	10.7	5387	13	US-10-001-873-22	Sequence 22, Appli
C	18	39.4	10.6	1389	14	US-10-156-761-5431	Sequence 5431, Ap
C	19	39.2	10.6	106664	14	US-10-175-523-97	Sequence 97, Appli
C	20	39	10.5	1341	14	US-10-156-761-4629	Sequence 4629, Ap
C	21	38.8	10.5	306	11	US-09-913-514-32	Sequence 32, Appli
C	22	38.8	10.5	434	10	US-09-960-352-2292	Sequence 2292, Ap
C	23	38.8	10.5	124884	11	US-09-913-514-1	Sequence 1, Appli
C	24	38.8	10.5	124884	12	US-10-288-823-76	Sequence 76, Appli
C	25	38.4	10.4	2946	14	US-10-128-714-7364	Sequence 7364, Ap
C	26	38.4	10.4	3000	14	US-10-128-714-2364	Sequence 2364, Ap
C	27	38.4	10.4	3169	14	US-10-128-714-1364	Sequence 1364, Ap
C	28	38.4	10.4	3169	14	US-10-128-714-6364	Sequence 6364, Ap
C	29	38.4	10.4	5168	14	US-10-128-714-564	Sequence 564, App
C	30	38.4	10.4	5169	14	US-10-128-714-5364	Sequence 5364, Ap
C	31	38	10.2	573	14	US-10-156-761-5782	Sequence 5782, Ap
C	32	38	10.2	861	14	US-10-156-761-3690	Sequence 3690, Ap
C	33	38	10.2	1429	14	US-10-128-846-13550	Sequence 13550, A
C	34	38	10.2	1885	14	US-10-106-698-420	Sequence 420, App
C	35	38	10.2	1223197	13	US-10-027-632-179264	Sequence 179264, Sequence 1, Appli
C	36	38	10.2	9025608	14	US-10-156-761-1	Sequence 6166, Ap
C	37	37.8	10.2	4548	14	US-10-156-761-6166	Sequence 143, App
C	38	37.6	10.1	1235	9	US-09-925-301-143	Sequence 3570, Ap
C	39	37.6	10.1	1869	14	US-10-151-832-4	Sequence 4, Appli
C	40	37.4	10.1	1294	14	US-10-151-832-4	Sequence 5780, Ap
C	41	37.2	10.0	477	14	US-10-037-270-1085	Sequence 1085, Ap
C	42	37.2	10.0	2817	14	US-10-037-270-1085	Sequence 20733, A
C	43	36.8	9.9	575	9	US-09-864-761-20733	Sequence 5073, Ap
C	44	36.8	9.9	1560	14	US-10-156-761-5073	Sequence 36, Appli
C	45	36.8	9.9	1886	11	US-09-373-658-36	

ALIGNMENTS

RESULT 1  
US-09-849-729-1  
; Sequence 1, Application US/09849729  
; Publication No. US20030165540A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jen-Kuei  
; APPLICANT: Lewis, Samantha  
; APPLICANT: Batz, Hans-Georg  
; APPLICANT: Ramaswamy, Iatcha  
; APPLICANT: Bohenzky, Roy  
; APPLICANT: Lin, Yu-Huei  
; APPLICANT: Montiel, Janine  
; APPLICANT: Chen, Benjamin  
; TITLE OF INVENTION: Sentinel Virus II  
; FILE REFERENCE: RDID 0070  
; CURRENT APPLICATION NUMBER: US/09/849,729  
; PRIOR FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: US 60/202271  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 371  
; TYPE: DNA  
; ORGANISM: Sentinel Virus II  
; US-09-849-729-1

Query Match  
Best Local Similarity 99.5%; Score 369; DB 12; Length 371;  
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGGAACGCTTGGCTCGATCGACGAAGACGGGATTCGA 60  
DB 1 GATCGGAACGCTTGGCTCGATCGACGAAGACGGGATTCGA 60  
QY 61 CGACGACGACGACGATTCGATGGAAGATGGAACGCGACGCGGCGGAGCG 120  
DB 61 CGACGACGACGACGATTCGATGGAAGATGGAACGCGACGCGGCGGAGCG 120

[illegible]

```

RESULT 2
US-09-849-729-2/c
; Sequence 2, Application US/09849729
; Publication No. US20030165540A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jen-Kuei
; APPLICANT: Lewis, Samantha
; APPLICANT: Batz, Hans-Georg
; APPLICANT: Ramaswamy, Latha
; APPLICANT: Bohenzky, Roy
; APPLICANT: Lin, Yu-Huei
; APPLICANT: Montiel, Janine
; APPLICANT: Chen, Benjamin
; TITLE OF INVENTION: Sentinel Virus II
; FILE REFERENCE: RDID 0070
; CURRENT APPLICATION NUMBER: US/09/849,729
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/202271
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Sentinel Virus II
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (372)..(372)
; OTHER INFORMATION: unknown: can be a, t, g, c or there could be no nucleotide found
;
US-09-849-729-2

```

Query Match	Similarity	Score	366.6	DB 12	Length 372
Best Local	Similarity	99.5%	Pred. No. 1.4e-104		
Matches 369	Conservative	0	Mismatches 2	Indels 0	Gaps 0
Qy	1	GATCMGAAACGTTSGCTGCGTGCATGCGAAGACGGSGVTGAAGCGGCGATTGA	60		
Db	371	GATCMGAAACGTTSGCTGCGTGCATGCGAAGACGGSGVTGAAGCGGCGATTGA	312		
Qy	61	CGACGACGACGATTCGATGTAAGTTGGACCGGCGAGTCCTTGCGCGGCGAGCG	120		
Db	311	CGACGACGACGATTCGATGTAAGTTGGACCGGCGAGTCCTTGCGCGGCGAGCG	252		
Qy	121	CGAGAACCAAGACGACGAGGACGAGCGGCTACGGCGCATCCGTTCTTCTTGAGCG	180		
Db	251	CGAGAACCAAGACGACGAGGACGAGCGGCTACGGCGCATCCGTTCTTCTTGAGCG	192		
Qy	181	GGATTATGACACTCCGCGATTCGTCGATCCGTAAGTTTCGCTTCGACAGCCTGATTCAG	240		
Db	191	GGATTATGACACTCCGCGATTCGTCGATCCGTAAGTTTCGCTTCGACAGCCTGATTCAG	132		

Qy	241	GCGCGTTTGGCCACCTGTTTGGACGCTGGAAACCGACGGCTGGACCATGCTAGTT	300
Db	131	GCGCGTTTGGCCACCTGCTTTGGACGCTGGAAACCGACGGCTGGACCATGCTAGTT	72
Qy	301	TCCTGCTTGAAGGTGAATTTCCGACCCCGTTGGACTTGTCTCTCTCTCTCTACCC	360
Db	71	TCCTGCTTGAAGGTGAATTTCCGACCCCGTTGGACTTGTCTCTCTCTCTCTCTACCC	12
Qy	361	CGACGACGATC	371
Db	11	CGACGACGATC	1

```

1      RESULT 3
2      US-09-790-399-7
3      ; Sequence 7, Application US/09790399
4      ; Patent No. US20020038000A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Gold, Larry
7      ; APPLICANT: Tuerk, Craig
8      ; APPLICANT: Pribnow, David
9      ; APPLICANT: Smith, Jonathan D.
10     ; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
11     ; FILE REFERENCE: NEX02/C1-C0N2
12     ; CURRENT APPLICATION NUMBER: US/09/790,399
13     ; CURRENT FILING DATE: 2001-02-22
14     ; PRIOR APPLICATION NUMBER: 09/197,649
15     ; PRIOR FILING DATE: 1998-11-23
16     ; PRIOR APPLICATION NUMBER: 07/829,461
17     ; PRIOR FILING DATE: 1992-01-31
18     ; PRIOR APPLICATION NUMBER: 07/739,055
19     ; PRIOR FILING DATE: 1991-08-01
20     ; PRIOR APPLICATION NUMBER: 07/561,968
21     ; PRIOR FILING DATE: 1990-08-02
22     ; NUMBER OF SEQ ID NOS: 26
23     ; SOFTWARE: PatentIn Ver. 2.0
24     ; SEQ ID NO 7
25     ; LENGTH: 390
26     ; TYPE: DNA
27     ; ORGANISM: Artificial Sequence
28     ; FEATURE:
29     ; OTHER INFORMATION: Description of Artificial Sequence: Sequence
30     ; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
31     ; OTHER INFORMATION: fragments having NcoI restriction sites.
32     ; US-09-790-399-7

```

```

Query Match      12.1%   Score 45;   DB 9;   Length 390;
Best Local Similarity 57.8%;   Pred. No. 0 0003;
Matches      78;   Conservative      1;   Mismatches      56;   Indels      0;   Gaps

```

QY

21 GGTCATGTCAGAAAGACGGGWTGAAGCGGACGGATTGACGACGACGACGATTTGCCA 80

DB 1 GGGCCATGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60

QY 81 TGAAGATGGGACCGCTCGACGTCCTTTGGCGGGCGGACGCCGAGAACCAAGACGAGG 14

DB 61 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAG 12

QY 141 ACGAGACGCTTACG 155

DB 121 ACGACGACGACGAG 135

RESULT 4

US-10-294-804-1

; Sequence 1, Application US/10294804

; Publication No. US20030133948A1

; GENERAL INFORMATION:

; APPLICANT: Robertson, Erly A.

; APPLICANT: Cotter, Murray S.

; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA

; TITLE OF INVENTION: to Genomic Host DNA

; FILE REFERENCE: UM-03778





```
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA-011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7738
LENGTH: 1437
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1437)
US-09-815-242-7738
```

```
Query Match      11.3%; Score 41.8; DB 9; Length 1437;
Best Local Similarity 52.9%; Pred. No. 0.0039;
Matches 82; Conservative 3; Mismatches 70; Indels 0; Gaps 0;
```

```
QY      12 GYTTSGCTGGTGCATGCAAGAGAGCGGWTGAAGCGGAGATTGACGACGACG 71
      1274 GCGGGCGTGGGCGGATGACGAGAGCGGCGGATGCTTCGATCGGCGGATG 1215
DB      72 ACATTGCGATGAAGATGGGACCGCTGACGCTCTTGGCGGCGGAGCGGAGAACCAAG 131
      1214 ACCTTGACCGAGCGGCTGTGTCATTGGGACGACCATCGCGACCGCTGGCGGAGCGG 1155
QY      132 ACGACGAGAGAGAGAGCGTCTACGCGGCGGATCGT 166
DB      1154 AAGGTGCGGACGTTGACTTGACGCCCTCGGCTT 1120
```

```
RESULT 12
US-10-156-761-2598
Sequence 2598, Application US/10156761
Publication No. US20030119018A1
```

```
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
```

```
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 2598
LENGTH: 498
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(498)
US-10-156-761-2598
```

```
Query Match      11.2%; Score 41.6; DB 14; Length 498;
Best Local Similarity 51.8%; Pred. No. 0.0036;
Matches 86; Conservative 3; Mismatches 77; Indels 0; Gaps 0;
```

```
QY      11 CGYTTSGCTGGTGCATGCAAGAGAGCGGWTGAAGCGGAGATTGACGACGAC 70
      265 CATGTCGAGGGGAGACCTGACCCGGTCCGAGAGCGCTGACCGCATGTGGAAGTCGAG 324
QY      71 GACATTGCGATGAAGATGGGACCGCYAGCGTCTTGGCGGCGGAGCGGAGAACCA 130
      325 TACGTGTGATGACCGCGGCTCTGTCGACATCTTCCGAGATGTCTGCGGAGAGACGAC 384
DB      131 GACGACGAGAGAGAGCGTCTACGCGGCGCATCGTCTTCTCTG 176
      385 GACCACTGTGAGCGTCATCAACGACATCCGGGCGCTGCCG 430
```

```
RESULT 13
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
```

```
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
```

```
US-10-156-761-1
```

```
Query Match      11.2%; Score 41.6; DB 14; Length 9025608;
Best Local Similarity 51.8%; Pred. No. 0.029;
Matches 86; Conservative 3; Mismatches 77; Indels 0; Gaps 0;
```

```
QY      11 CGYTTSGCTGGTGCATGCAAGAGAGCGGWTGAAGCGGAGATTGACGACGAC 70
      3210013 CATGTCGAGGGGAGACCTGACCCGGTCCGAGAGCGCTGACCGCATGTGGAAGTCGAG 3210072
QY      71 GACATTGCGATGAAGATGGGACCGCYAGCGTCTTGGCGGCGGAGCGGAGAACCA 130
      3210073 TACGTGTGATGACCGCGGCTCTGTCGACATCTTCCGAGATGTCTGCGGAGAGACGAC 3210132
DB      131 GACGACGAGAGAGAGCGTCTACGCGGCGCATCGTCTTCTCTG 176
```

Db 3210133 GACCACTGCTGACATCAACAAACGATCCGGCCCTGCGCC 3210178

RESULT 14  
US-09-913-514-31  
Sequence 31, Application US/09913514  
Publication No. US20030082210A1  
GENERAL INFORMATION:

APPLICANT: GOMI, Yasuyuki  
APPLICANT: SUNAMACHI, Hiroki  
APPLICANT: TAKAHASHI, Michiaki  
APPLICANT: YAMAMOTO, Koichi

TITLE OF INVENTION: Method for Quality Control of an Attenuated Varicella Live Vaccine

FILE REFERENCE: 0216-0454P  
CURRENT APPLICATION NUMBER: US/09/913, 514  
CURRENT FILING DATE: 2001-12-07

PRIOR APPLICATION NUMBER: PCT/JP01/00678

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: JP 2000-62734

PRIOR FILING DATE: 2000-01-31

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.1

SEQ ID NO 31

LENGTH: 228

TYPE: DNA

ORGANISM: Varicella virus

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)..(228)

OTHER INFORMATION: Parental Oka strain

US-09-913-514-31

Query Match 10.9%; Score 40.4; DB 11; Length 228;  
Best Local Similarity 58.2%; Pred. No. 0.0073;  
Matches 71; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 28 GCAGAAAGACGGGWTGAAGCGCGGATTTGACGACGACATTGGCATGAAGA 87  
Db 76 GGAGAGAGAGAGAGCGGAGGAGGAGCGGATTGACGACGAGGAGAGCGCGGAGGAGGA 135

QY 88 TGGGACCGCYGACCTCTTGCGGGGCGGAGCGCGAGAACCAAGACGACGAGAGGA 147  
Db 136 CGCGGAGAGAGAGCGCGGCGGAGAGAGACCGCATCGACGAGGAGAGAGCGGAGAGGA 195

QY 148 CG 149  
Db 196 CG 197

RESULT 15  
US-09-771-208-20/c

Sequence 20, Application US/09771208

Patent No. US20020155564A1

GENERAL INFORMATION:

APPLICANT: MEDRANO, JUAN

APPLICANT: BRADFORD, ERIC

APPLICANT: HORVAT, SIMON

TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE

FILE REFERENCE: 407T-923710US

CURRENT APPLICATION NUMBER: US/09/771,208

CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: US 08/999,477

PRIOR FILING DATE: 1997-12-29

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn version 3.0

SEQ ID NO 20

LENGTH: 659158

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (123459)..(123478)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (602466)..(602485)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (546988)..(547017)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (494715)..(494814)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (390986)..(391005)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (346860)..(346883)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (317174)..(317193)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (280353)..(280373)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (271829)..(271848)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (183872)..(183891)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (170625)..(170645)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

OTHER INFORMATION: n is unidentified a, c, g, or t

NAME/KEY: misc\_feature

LOCATION: (132680)..(132700)

Search completed: September 15, 2003, 01:28:07  
Job time : 192 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 23:14:29 ; Search time 1967 Seconds  
(without alignments)  
4584.118 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371  
Sequence: 1 gatcmgaacgyltsgtc.....ctctaccgcagcagc 371

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estmu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gse\_hum:\*  
18: em\_gse\_hnv:\*  
19: em\_gse\_pln:\*  
20: em\_gse\_vrt:\*  
21: em\_gse\_fun:\*  
22: em\_gse\_mam:\*  
23: em\_gse\_mus:\*  
24: em\_gse\_pro:\*  
25: em\_gse\_rtd:\*  
26: em\_gse\_phg:\*  
27: em\_gse\_vrl:\*  
28: gb\_gse1:\*  
29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	48.4	13.0	879	29	CNS03151	AL223102 Tetradon
2	47.2	12.7	487	10	BE494661	BE494661 WHE1257_G
3	47.2	12.7	490	10	BE494611	BE494611 WHE1255_C
4	47.2	12.7	557	10	BF145396	BF145396 WHE1841-1

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
CNS03151	CNS03151	Tetradon nigroviridis genome survey sequence T7 end of clone 204608 of library G from Tetradon nigroviridis, genomic survey sequence.	AL223102	AL223102.1	GI:7881921	GSS; genome survey sequence.	Tetradon nigroviridis	Roest Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brotier, P., Quetier, F., Sautin, W. and Weissenbach, J.	Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence	Nat. Genet. 25 (2), 235-238 (2000)	20296633	10835645	Roest Crollius, H., Jallion, O., Dasilva, C., Ozouf-Costez, C., Fizames, C., Fischer, C., Bouneau, L., Billaule, A., Quetier, F.,		

## ALIGNMENTS

C	5	45.4	12.2	816	10	BE621513
6	45	12.1	464	10	BE799902	BE799902 601588045
7	44.6	12.0	485	14	CA719534	CA719534 wkmc2n.pk0
8	44.4	12.0	1228	29	BZ575028	BZ575028 msn2_4267
9	43.8	11.8	444	14	CD056167	CD056167 H011G01S
10	43.8	11.8	449	14	CB085540	CB085540 hg11g07.9
11	43.8	11.8	470	14	CA006622	CA006622 HU05D13r
12	43.8	11.8	514	13	BQ464793	BQ464793 HU01114T
13	43.8	11.8	530	14	CA029558	CA029558 HZ65J01r
14	43.8	11.8	531	14	CB859920	CB859920 H112A05w
15	43.8	11.8	923	13	BU771970	BU771970 SJERLA02
16	43.8	11.8	1942	13	BU771792	BU771792 SJERLA08
17	43.4	11.7	631	28	BZ117262	BZ117262 CH230-255
18	43.4	11.7	1004	13	BU772046	BU772046 SJERLGO7
19	43.4	11.7	1025	13	BU770240	BU770240 SJERCTH03
20	43.4	11.7	1124	10	BF314946	BF314946 601899453
21	43.4	11.7	1315	9	AV752548	AV752548 AV752548
22	43.2	11.6	1091	13	BU766419	BU766419 SJEREA02
23	43	11.6	2003	10	BG330570	BG330570 602430456
24	42.8	11.5	301	10	BG605592	BG605592 WHE2239_F
25	42.6	11.5	324	10	BF986279	BF986279 QV4-GN014
26	42.6	11.5	492	12	BM377307	BM377307 EBem05_SQ
27	42.6	11.5	552	9	AL821572	AL821572 AL821572
28	42.6	11.5	731	13	BU766448	BU766448 SJEREC07
29	42.6	11.5	2162	13	BU766490	BU766490 SJERAG03
30	42.4	11.4	365	29	I2644X	AL160094 Leishmani
31	42.4	11.4	400	14	CA020093	CA020093 HVL4F04r
32	42.4	11.4	507	10	BE602842	BE602842 HVSMEH010
33	42.4	11.4	560	10	BE515727	BE515727 WHE0602_A
34	42.4	11.4	585	10	BF255377	BF255377 HVSMEF000
35	42.4	11.4	799	10	BE412116	BE412116 JUI002_G0
36	42.4	11.4	837	10	BG344165	BG344165 HVSMEG000
37	42.4	11.4	846	14	CB644680	CB644680 OSJEB06A
38	42.2	11.4	234	13	BU771897	BU771897 SJERXB12
39	42.2	11.4	504	10	BS274381	BS274381 WHE2228_A
40	42.2	11.4	560	9	AV430031	AV430031 AV430031
41	42.2	11.4	566	10	BF292426	BF292426 WHE2213_F
42	42.2	11.4	575	14	CA599404	CA599404 waw1c.pk0
43	42.2	11.4	588	10	BE060167	BE060167 HVSMEG001
44	42.2	11.4	589	14	CA599424	CA599424 waw1c.pk0
45	42.2	11.4	663	14	CD233070	CD233070 SSL_11_H0

TITLE	Saurin, W., Bernot, A. and Weissenbach, J.		
JOURNAL	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
MEDLINE	Genome Res. 10 (7), 939-949 (2000)		
PUBMED	20359837		
REFERENCE	10899143		
AUTHORS	3 (bases 1 to 879)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
COMMENT	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
FEATURES	- Web : www.genoscope.cns.fr) This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> .		
source	Location/Qualifiers		
BASE COUNT	1. 879		
ORIGIN	/organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone="204G08" /clone_lib="G" /note="Genoscope sequence ID : CONG204BD04LP1-end : T7"		
Query Match	13.0%; Score 48.4; DB 29; Length 879;		
Best Local Similarity	58.5%; Pred. 0.095;		
Matches	76; Conservative 4; Mismatches 50; Indels 0; Gaps 0;		
QY	26 ATCGAAGAAGACGGGTGAAGCCGACCGGATTGACGACGACGACATTCGATGAA 85		
DB	353 AAGGAGGAGGACCAAGGAAGAGAGAGAGACGAGCAAGGAGAGAGCAAGAA 412		
QY	86 GATGGAGACCGCGACGCTCTTGCGCGGCGAGCCGGAACCAAGACGACGAGAG 145		
DB	413 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 472		
QY	146 GACGCTTACG 155		
DB	473 GAGGACCAAG 482		
RESULT 2	BEA94661 487 bp mRNA linear EST 02-AUG-2000		
LOCUS	BEA94661		
DEFINITION	WHE1257_G04_N07ZS Secale cereale anther cDNA library Secale cereale		
ACCESSION	BEA94661		
VERSION	BEA94661.1 GI:9661254		
KEYWORDS	EST.		
SOURCE	Secale cereale (rye)		
ORGANISM	Secale cereale		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spemmatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poidea		
AUTHORS	1 (bases 1 to 487)		
TITLE	Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton		
JOURNAL	R.D., Gustafson, J.P., Han, P.S., Hsieh, C.C., Kang, Y., Lazo, G.R.,		
COMMENT	Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C. The structure and function of the expressed portion of the wheat genomes - Another cDNA library from rye Unpublished Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderson@w.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20		

FEATURES	Seq primer: Stratiogene SK primer.
Source	Location/Qualifiers
	1..487
	/organism="Secale cereale"
	/mol_type="mRNA"
	/cultivar="Blanco"
	/db_xref="taxon:4550"
	/clone="WHE1257.G04.N07"
	/tissue_type="Anther"
	/dev_stage="Adult plant before anthesis"
	/lab_host="E. coli SOLR"
	/clone_lib="Secale cereale anther cDNA library"
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse. Anthers were harvested and pooled from early meiosis to late meiosis. The tissue, total RNA, and poly(A) RNA were prepared (Butler, Ross and Gustafson) at University of Missouri, Columbia. A cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the TV Close lab (Choi, Close, Penton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT	120 a 161 c 138 g 68 t
ORIGIN	
Query Match	12.7%; Score 47.2; DB 10; Length 487;
Best Local Similarity	60.3%; Pred. No. 0.17; 49; Indels 0; Gaps 0;
Matches	76; Conservative 1; Mismatches
Oy	30 AGAAGACGGGWTGAAGCGCGAGCGGATTCAGCAGCAGCAGCATTCGATGAAAGATG 89
Db	349 ACAAGGACTACGACAAAGGTGACGACAGCAGCAGCATGATGACAAAGATGATG 408
Oy	90 GGACCGCYGACGTCCTTTGGCGGGCGGAGCGCGAACCAGACGACGAGAGCAGAGACG 149
Db	409 ACGACGACGACCAAGCAGCAGCAGCAGGAGAGCAGCAGCAGCAGCAGATGACG 468
Oy	150 TCTACG 155
Db	469 ACGACG 474
RESULT 3	
BE494611	490 bp mRNA linear EST 02-AUG-2000
LOCUS	WHE1255_C07_E13S Secale cereale anther cDNA library Secale cereale
DEFINITION	CDNA clone WHE1255_C07_E13, mRNA sequence.
ACCESSION	BE494611
VERSION	BE494611.1 GI:9661204
KEYWORDS	EST.
SOURCE	Secale cereale (rye)
ORGANISM	Secale cereale Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae i (bases 1 to 490) 1 (bases 1 to 490) Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Penton,R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C. The structure and function of the expressed portion of the wheat genomes - Anther cDNA library from rye
REFERENCE	Unpublished
AUTHORS	Contact: Olin Anderson
JOURNAL	US Department of Agriculture, Agriculture Research Service, Pacific
COMMENT	West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanders@pw.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stratiogene SK primer.

FEATURES  
source

Location/Qualifiers  
1. .490  
/organism="Secale cereale"  
/mol\_type="mRNA"  
/cultivar="Blanco"  
/db\_xref="taxon:4550"  
/clone="WHE1255\_C07\_E13"  
/tissue\_type="Anther"  
/dev\_stage="Adult plant before anthesis"  
/lab\_host="E. coli SOLR"  
/clone\_lib="Secale cereale anther cDNA library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in the  
greenhouse. Anthers were harvested and pooled from early  
meiosis to late meiosis. The tissue, total RNA, and  
poly(A) RNA were prepared (Butler, Ross and Gustafson) at  
University of Missouri, Columbia. A cDNA library was  
made, and the cDNA clones were in vivo excised to give  
phagescript phagemids in the TJ Close lab (Choi, Close,  
Fenton) at the University of California, Riverside.  
Plasmid DNA preparations and DNA sequencing were performed  
in the OD Anderson lab (all other authors)."

BASE COUNT 121 a 163 c 138 g 68 t

BASE COUNT 137 a 175 c 164 g 80 t 1 others

Query Match 12.7%; Score 47.2; DB 10; Length 490;  
Best Local Similarity 60.3%; Pred. No. 0.17;  
Matches 76; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

Query Match 12.7%; Score 47.2; DB 10; Length 557;  
Best Local Similarity 60.3%; Pred. No. 0.17;  
Matches 76; Conservative 1; Mismatches 49; Indels 0; Gaps 0;

QY 30 AGAAGACCGGWTGAAGCGCGACGCGATTGACGACGACGACGATTCGATGAAGAAG 89  
DB 351 ACAAGGACTACGACAAAGGTGACGACGACGACGACGACGATGACGATGACGATGACG 410  
QY 90 GGACGCGYACGCTCTTGGCGGGGCGGAGCGGAGAACCAAGACGACGACGACGAGAGC 149  
DB 411 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 470  
QY 150 TCTACG 155  
DB 471 ACGACG 476

QY 30 AGAAGACCGGWTGAAGCGCGACGCGATTGACGACGACGACGACGATTCGATGAAGAAG 89  
DB 323 ACAAGGACTACGACAAAGGTGACGACGACGACGACGACGATGACGATGACGATGACG 382  
QY 90 GGACGCGYACGCTCTTGGCGGGGCGGAGCGGAGAACCAAGACGACGACGACGAGAGC 149  
DB 383 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 442  
QY 150 TCTACG 155  
DB 443 ACGACG 448

RESULT 4  
BF145396 557 bp mRNA linear EST 26-OCT-2000  
LOCUS WHE1841-1844\_B21\_B21S2 Secale cereale anther cDNA library Secale  
DEFINITION cereale cDNA clone WHE1841-1844\_B21\_B21, mRNA sequence.  
ACCESSION BF145396  
VERSION BF145396.1 GI:11026827  
KEYWORDS EST.  
SOURCE Secale cereale (rye)  
ORGANISM Secale cereale

RESULT 5  
BF621513 816 bp mRNA linear EST 18-OCT-2001  
LOCUS HVSMEa0011D21f Hordeum vulgare seedling shoot EST library  
DEFINITION HVSMEa0011D21f, mRNA sequence.  
ACCESSION BF621513  
VERSION BF621513.3 GI:16256502  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE  
AUTHORS Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Gustafson, J.P., Han, P.S., Heia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C.  
TITLE The structure and function of the expressed portion of the wheat genomes - Anther cDNA library from rye  
JOURNAL Unpublished  
COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: StrataGene SK primer.  
Location/Qualifiers

REFERENCE  
AUTHORS Wing, R., Close, T.J., Kleinof, A., Wise, R., Begum, D., Fritsch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R., Choi, D.W., Fenton, R.D. and Mann, D.  
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex cold-stressed seedling shoot cDNA library  
JOURNAL Unpublished  
COMMENT On Dec 18, 2000 this sequence version replaced gi:13083392.  
Contact: Wing RA  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu  
Total hg bases = 552  
Seq primer: AATTAACCTCACTAAAGG

FEATURES

Location/Qualifiers

High quality sequence stop: 641.  
 FEATURES  
 Location/Qualifiers

source

1. 816  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="rRNA"  
 /cultivar="Morex"  
 /db\_xref="taxon:112509"  
 /clone="HVSMA0011D21f"  
 /rname\_type="Seedling shoot"  
 /lab\_host="TJC121"  
 /clone\_id="Hordeum vulgare seedling shoot EST library  
 HVCDA0001 (Cold stress)"  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 Seeds were surface sterilized then germinated under axenic  
 conditions in the dark at room temperature on filter paper  
 with water, nystatin and cefotaxime in covered  
 crystallization dishes. Five-day old seedlings were  
 incubated at 50C for 2 days. Shoots were then harvested,  
 local RNA was prepared, poly(A) RNA was purified, one  
 primary unamplified cDNA library was made, and 600000 pfu  
 were in vivo excised to give Bluescript SK(-) cDNA  
 phagemids. These steps were performed in the U.C. Close  
 laboratory at the University of California, Riverside  
 (Choi, Close, Fenton). Phagemids were plated and picked at  
 the Clemson University Genomics Institute (CUGI) (Begum,  
 Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations  
 , DNA sequencing and sequence analysis were performed at  
 CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main  
 ). The sequence has been trimmed to remove vector sequence  
 and contains a minimum of 100 bases of phred value 20 or  
 above. For more details on library preparation and  
 sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders Also  
 see Close TJ, Wing R, Kleinbols A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"

BASE COUNT  
 ORIGIN  
 141 a 270 c 271 g 134 t

Query Match 12.2%; Score 45.4; DB 10; Length 816;  
 Best Local Similarity 61.9%; Pred. No. 0.51;  
 Matches 70; Conservative 1; Mismatches 42; Indels 0; Gaps 0;

OY 52 CGGATTGACGACGACGACGACGATTCGATGAAGATGGACCCGCGACGCTCTTGGCG 111  
 DB 494 CGTCTCCACGACGCTCCGCGCCGAGCGCGGTGCGACGCTCGGTGCGCGGCGC 435  
 OY 112 GCGGAGCGGAGAACGACGACGACGACGACGCTTACGCGCGCATCC 164  
 DB 434 GCCACTGCTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 382

RESULT 6  
 BE799902 464 bp mRNA linear EST 20-SEP-2000  
 LOCUS 601588045F1 NIH\_MGC\_7 Homo sapiens cDNA IMAGE:3942251 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE799902  
 VERSION BE799902.1 GI:10221100  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 464)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov  
 Plate: LLCM795 row: j column: 12  
 High quality sequence start: 69  
 High quality sequence stop: 215.  
 Location/Qualifiers

FEATURES  
 source

1. 464  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3942251"  
 /rname\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="PH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC-7"  
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GCGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT  
 ORIGIN  
 106 a 125 c 145 g 88 t

Query Match 12.1%; Score 45; DB 10; Length 464;  
 Best Local Similarity 55.9%; Pred. No. 0.58;  
 Matches 81; Conservative 2; Mismatches 62; Indels 0; Gaps 0;

OY 11 CGTTGGCTCGGTGACATGACGAAAGACGCGGWTGAAGCGGATTCGACGACGAC 70  
 DB 32 CGGTAAATATCTCCGCGCCCGACGAGGACGACGACGACGACGACGACGACGAC 91  
 OY 71 GACATTGCGATGAAAGATGGGACCGCGACGCTTGGCGGCGGAGCGGAAACCA 130  
 DB 92 GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 151  
 OY 131 GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 155  
 DB 152 GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 176

RESULT 7  
 CA719534 485 bp mRNA linear EST 26-NOV-2002  
 LOCUS wkm2n.pk004.g21 wkm2n Triticum aestivum cDNA clone wkm2n.pk004.g21  
 DEFINITION 5' end, mRNA sequence.  
 ACCESSION CA719534  
 VERSION CA719534.1 GI:25441327  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 ; Triticeae; Triticum.  
 REFERENCE 1 (bases 1 to 485)  
 AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,  
 Miao,G., Caraher,N. and Hanafey,M.K.  
 TITLE Dupont wheat cDNA sequence  
 JOURNAL Unpublished  
 COMMENT Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. Dupont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13  
 Location/Qualifiers

FEATURES

```

source
1. .485
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Spring wheat"
/db_xref="taxon:4565"
/clone="wkm2n.pk004.g21"
/issue_type="kernel"
/lab_host="DH10B"
/clone_lib="wkm2n"
/notes="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) kernel malted 175 hours
at 4 °C, normalized"
BASE COUNT      99 a      147 c      172 g      65 t      2 others
ORIGIN

Query Match
Best Local Similarity 52.0%; Score 44.6; DB 14; Length 485;
Matches 89; Conservative 4; Mismatches 78; Indels 0; Gaps 0;

QY 12 GTTSGCTCGTGATGACAGAGACGGGWTGAAGCGGACGAGATTGACGACGACG 71
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 92 GCATGACTGCGCGCGGTGAGAGACGACGAGACGACCATGCGCGGACGCGCGG 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 72 ACATTGCGATGAAGATGGGACCGCTGACGCTCTTGGCGGCGGACCGAGACCAAG 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 152 ACCCGCGGAGTGAAGAGTGGCGGACGCTCCGCGCGCTGTGTGAGGCGGACCGCG 211
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 132 ACGACGAGACGAGACGCTACGCGGACGCTTCTCTCTGAGCGGCG 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 212 CAGCCAGAGAGAGAGACGACTTATCTGCGCGCTTCTGCGGCTCGAG 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 8
BZ575028      1228 bp      DNA      linear      GSS 17-DEC-2002
LOCUS      msh2_4267.v2 msh Pseudomonas aeruginosa genomic clone msh2_4267,
DEFINITION      genomic survey sequence.
ACCESSION      BZ575028
VERSION      BZ575028.1 GI:27210089
KEYWORDS      GSS.
SOURCE      Pseudomonas aeruginosa
ORGANISM      Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 1228)
AUTHORS      Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
              Burns,U.L., Kaul,R. and Olsen,M.V.
              Whole-Genome-Sequence variation among multiple isolates of
              Pseudomonas aeruginosa library
              J. Bacteriol., (2002) In press
              Contact: Chris K. Raymond
              Genome Center
              University of Washington
              Box 352145, Seattle, WA 98105-2145, USA
              Tel: 2062216954
              Fax: 2066857244
              Email: craymond@u.washington.edu
              Class: shotgun.
FEATURES
      source
      Location/Qualifiers
      1..1228
      /organism="Pseudomonas aeruginosa"
      /mol_type="genomic DNA"
      /strain="MSH"
      /db_xref="taxon:287"
      /clone="msh2_4267"
      /clone_lib="msh"
      /notes="Environmental isolate. Whole genomic shotgun
      library."
BASE COUNT      250 a      345 c      275 g      357 t      1 others
ORIGIN

Query Match
Best Local Similarity 12.0%; Score 44.4; DB 29; Length 1228;
Matches 50.5%; Pred. No. 0.97;

```

```

Matches 99; Conservative 3; Mismatches 94; Indels 0; Gaps 0;

QY 12 GTTSGCTCGTGATGACAGAGACGGGWTGAAGCGGACGAGATTGACGACGACG 71
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 210 GCGGCGCTGGGCGCATACGATGAGACCGCAGACGCGGTCTGTCATCGGAGAG 269
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 72 ACATTGCGATGAAGATGGGACCGCTGACGCTCTTGGCGGCGGACCGGAGAACCAAG 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 270 ACCTTGACAGGCGCGGTGTGCTGTGGCAGCATGCGGACCGCTGCGCGGCGAACGCG 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 132 ACGACGAGACGAGACGCTTACGCGCGGATCCGTTCTCTTACGCGGATTTGA 191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 330 AAGTGCCGACGTTGACTTGACGCCCTTGAGGGGTCTGCTGCTTTGCGGAC 389
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 192 CCTCGCATTTGCTGAT 207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 390 CAGGCGATTTCCGAT 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 9
CD056167/c      444 bp      mRNA      linear      EST 05-JUN-2003
LOCUS      H011G01S HO Hordeum vulgare cDNA clone H011G01 5-PRIME, mRNA
DEFINITION      sequence.
ACCESSION      CD056167
VERSION      CD056167.1 GI:30598733
KEYWORDS      EST.
SOURCE      Hordeum vulgare
ORGANISM      Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE      1 (bases 1 to 444)
AUTHORS      Zierold,U. and Schweizer,P.
              Barley ESTs from pathogen-attacked leaf epidermis
              Unpublished
              Contact: Patrick Schweizer
              Transcription Analysis, Cytogenetics Department
              Institute of Plant Genetics and Crop Plant Research (IPK)
              Corrensstr. 3, D-06466 Gatersleben, Germany
              Tel: 0049 (0)39482-5660
              Fax: 0049 (0)39482-5595
              Email: schweiz@ipk-gatersleben.de
              Insert Length: 444 Std Error: 0.00
              Plate: 11 row: G column: 1
              Seq primer: SK.
FEATURES
      source
      Location/Qualifiers
      1..444
      /organism="Hordeum vulgare"
      /mol_type="mRNA"
      /cultivar="Ingrid BC mlo-5"
      /db_xref="GABI:707130"
      /db_xref="taxon:4513"
      /clone="H011G01"
      /tissue_type="leaf epidermis, 6 h and 24 h post
      inoculation with Blumeria graminis"
      /dev_stage="7 d after germination"
      /lab_host="XLI0-Gold"
      /clone_lib="HO"
      /notes="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
      cDNA); Site 2: XhoI (3'-end of cDNA); Approximately 5 % of
      the clones correspond to cDNA from the fungi B. graminis
      hordei and tritici, respectively. Due to a cloning
      artefact caused by the kit, in most cases the EcoRI site
      is NOT present, as well as the EcoRI adapter used for
      cloning. To excise the insert, restriction sites upstream
      EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
      due to the cloning system used Blue/white selection for
      recombinants is not 100% reliable. Average insert size is
      1.2 kb"
BASE COUNT      68 a      143 c      173 g      60 t
ORIGIN

```

Query Match 11.8%; Score 43.8; DB 14; Length 444;  
 Best Local Similarity 61.1%; Pred. No. 1.1;  
 Matches 69; Conservative 1; Mismatches 43; Indels 0; Gaps 0;

Oy 52 CGGATTGACGACGACGACATTCGATGAAAGATGGACCGCTCTTGGCGG 111  
 Db 424 CGTGCTCCACGACGCTCGCGCCCGCGGGCTCGACGGCTCCGGTACCGTCCGCGCC 365

Oy 112 GCGGAGCGCGGACCAACGACGACGACGACGCTCTACCGCGCGATCC 164  
 Db 364 GCCACTGTGTGAGACCCAGTCCACGACGACGAGGTGTGTTCGCGCGCTCC 312

RESULT 10  
 CB085540 449 bp mRNA linear EST 27-JAN-2003  
 LOCUS hg11907.g1 Hedyotis centranthoides flower - Stage 2 (NYBG) Hedyotis  
 DEFINITION centranthoides cDNA clone hg11907, mRNA sequence.  
 ACCESSION CB085540  
 VERSION CB085540.1 GI:27909732  
 KEYWORDS EST.  
 SOURCE Hedyotis centranthoides  
 ORGANISM Hedyotis centranthoides  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamids; Gentianales; Rubiaceae; Rubioidae; SpERMACEAE  
 ; Hedyotis.  
 1 (bases 1 to 449)  
 Levesque, M.P., Twigg, R.W., Mole, T., Katari, M.S., Dedhia, N.N.,  
 O'Shaughnessy, A.L., Balija, V., Martienssen, R.A., McCombie, R.W.,  
 Bentley, P. and Stevenson, D.  
 Expressed tag sequences from Hedyotis centranthoides flower - Stage  
 2 (NYBG)  
 Unpublished  
 Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: hg11 row: 5 column: 07  
 Seg primer: -21M13Univrev  
 High quality sequence stop: 449.  
 Location/Qualifiers  
 1..449  
 /organism="Hedyotis centranthoides"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:219666"  
 /clone="hg11907"  
 /dev\_stage="pre-anthesis; Stage 2"  
 /clone\_id="Hedyotis centranthoides flower - Stage 2 (NYBG  
 )"  
 /note="Organ: flower; Vector: pBK-CMV; Site 1: XhoI;  
 Site 2: Eco RI; Date: Completed 12/18/01. Submitted to  
 CSHL 12/21/01 Library: Stratagene ZAP Express cDNA  
 Synthesis Kit. The library was size-fractionated to enrich  
 for large inserts. Sample: collected on the island of  
 Hawaii, Hawaii; NYBG herbarium voucher TM2563"

BASE COUNT 135 a 60 c 157 g 97 t

ORIGIN

Query Match 11.8%; Score 43.8; DB 14; Length 449;  
 Best Local Similarity 58.5%; Pred. No. 1.1;  
 Matches 72; Conservative 2; Mismatches 49; Indels 0; Gaps 0;

Oy 27 TGCAGAGGACGCGGATGAAAGCGGATGACGACGACGATTTGGGATGAAG 86  
 Db 91 TGGAGAGGAAATTAAGGCGGCTGACGAGTTAAAGCAAGAGAGTTGGGAGAG 150

Oy 87 ATGGAGCCGCGACGCTCTTGGCGGCGGACGCGGACGACGACGAGAGG 146  
 Db 151 AGCTGCTGTGATGACGTTGAAGATGAAGACGATGACGACGACGACGACG 210

Oy 147 ACG 149  
 Db 211 ACG 213

RESULT 11  
 CA006622/c 470 bp mRNA linear EST 23-OCT-2002  
 LOCUS HU05D13r HU Hordeum vulgare subsp. vulgare cDNA clone HU05D13  
 DEFINITION 5-PRIME, mRNA sequence.  
 ACCESSION CA006622  
 VERSION CA006622.1 GI:24283604  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 470)  
 Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and Graner  
 , A.  
 Barley ESTs from germinating seeds  
 Unpublished  
 Contact: Stein Nils  
 Molecular Markers Group, Department Genbank  
 Institute of Plant Genetics and Crop Plant Research (IPK)  
 Corrensstr. 3, 06466, Gatersleben, Germany  
 Tel: 039482-5522  
 Fax: 039482-5595  
 Email: stein@ipk-gatersleben.de  
 Insert Length: 470 Std Error: 0.00  
 Plate: 5 row: D column: 13  
 Seq primer: M13rev.  
 Location/Qualifiers  
 1..470  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="barke"  
 /db\_xref="GABI:264272"  
 /db\_xref="taxon:112509"  
 /clone="HU05D13"  
 /tissue\_type="germinating seeds"  
 /dev\_stage="germinating seeds, 16-48 h"  
 /lab\_host="XL10-Gold"  
 /clone\_id="HU"  
 /note="Vector: pBluescript SK+, Site 1: EcoRI (5'-end of  
 cDNA); Site 2: XhoI (3'-end of cDNA); germinating seeds,  
 16-48h. Due to a cloning artefact caused by the kit, in  
 most cases the EcoRI site is NOT present, as well as the  
 EcoRI adapter used for cloning. To excise the insert,  
 restriction sites upstream EcoRI should be used (e.g.  
 BamHI, SalI, PstI). NOTE: Also due to the cloning system  
 used Blue/white selection for recombinants is not 100%  
 reliable. Average insert size is 1.2 Kb"

BASE COUNT 69 a 166 c 167 g 68 t

ORIGIN

Query Match 11.8%; Score 43.8; DB 14; Length 470;  
 Best Local Similarity 61.1%; Pred. No. 1.2;  
 Matches 69; Conservative 1; Mismatches 43; Indels 0; Gaps 0;

Oy 52 CGGATTGACGACGACGACATTCGATGAAAGATGGACCGCTCTTGGCGG 111  
 Db 277 CGTGCTCCACGACGCTCGCGCCCGCGGGCTCGACGGCTCCGGTACCGTCCGCGCG 218

Oy 112 GCGGAGCGCGGACCAACGACGACGACGACGCTCTACCGCGCATCC 164  
 Db 217 GCCACTGTGTGAGACCCAGTCCACGACGACGAGGTGTGTTCGCGCGCTCC 165

RESULT 12  
 B0464793/c

[illegible]

Corrensstr. 3; 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@ipk-gatersleben.de

Insert Length: 531 Std Error: 0.00

Plate: 12 row: A column: 5

Seq primer: 77.

# FEATURES

## source

Location/Qualifiers

1..531

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="barke"

/db\_xref="GABI:557503"

/db\_xref="taxon:112509"

/clone="H112A05"

/tissue\_type="female inflorescences"

/dev\_stage="female inflorescences (approx. 3 mm in size)"

/lab\_host="XLI0-Gold"

/clone\_lib="Hi"

/note="Vector: pBluescript SK+; Site\_1: EcoRI (5'-end of cDNA); Site\_2: XhoI (3'-end of cDNA). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream of EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

## BASE COUNT

112 a 163 c 166 g 90 t

## ORIGIN

Query Match 11.8%; Score 43.8; DB 14; Length 531; Best Local Similarity 61.1%; Pred. No. 1.2; Matches 69; Conservative 1; Mismatches 43; Indels 0; Gaps 0;

QY 52 CGGATTGACGACGACGACGATTCGATGAAGATGGACCGCYGACGCTTCGCGG 111  
DB 383 CGTCTCCAGACGACGCTCCCGCCGCGCGGCTCGTCCGATGACGCTCGCGCGC 442  
QY 112 GCGGAGCGCGAGAACGACGACGACGACGACGCTCTAGCGCGCATCC 164  
DB 443 GCCACTGTGTGACGACCCCACTCCACGACGACGAGGTGTGTCCGCGCGCTCC 495

## RESULT 15

### LOCUS

BU771970 923 bp mRNA linear EST 12-FEB-2003

### DEFINITION

SJEBLA02 SJF Schistosoma japonicum cDNA, mRNA sequence.

### ACCESSION

BU771970.1 GI:28339286

### KEYWORDS

EST.

### SOURCE

Schistosoma japonicum

### ORGANISM

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeididae; Schistosomatidae; Schistosomatidae; Schistosoma.

### REFERENCE

1 (bases 1 to 923)

### AUTHORS

Hu, W., Yan, Q., Shen, D., Liu, F., Xu, X., Zhu, Z., Zhang, X., Wang, J., Sun, J., Xu, X., Wang, Z., Zeng, L., Rong, Y., Wu, X., Qu, J., Xu, Z., Huang, J., Ma, Y., Wang, S., Wang, Z., Xue, C., Feng, Z., Chen, Z. and Han, Z.

### TITLE

Expressed sequence tags from eggs of Schistosoma japonicum

### JOURNAL

Unpublished

### COMMENT

Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn.

### FEATURES

#### source

#### location/Qualifiers

1..923

#### organism

"Schistosoma japonicum"

#### mol\_type

"mRNA"

#### db\_xref

"taxon:6182"

/tissue\_type="whole egg"

/dev\_stage="egg"

/lab\_host="rabbits"

/clone\_lib="SJF"

303 a 405 c 179 g 34 t 2 others

## BASE COUNT

303 a 405 c 179 g 34 t 2 others

## ORIGIN

### Query Match

11.8%; Score 43.8; DB 13; Length 923; Best Local Similarity 56.8%; Pred. No. 1.3; Matches 75; Conservative 2; Mismatches 55; Indels 0; Gaps 0;

QY 24 GCATGCAAGAGACGGGWTGAAGCGGACGATGACGACGACGACGATTCGATGA 83  
DB 10 GCAAGNAACATATCGGATGACGATGACGACGACGACGACGACGACGACGACG 69  
QY 84 AAGATGGACCGCYGACGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143  
DB 70 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 129  
QY 144 AGGACGCTTACG 155  
DB 130 ACGACGACGACG 141

Search completed: September 15, 2003, 00:33:36  
Job time : 1971 secs



Tue Sep 16 09:06:26 2003

US-09-849-729-1.rmpm

Schenker L.  
09/18/97 729 Page 1  
Seq-10 1 w/ Inters

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 23:20:05 ; Search time 2867 Seconds  
(without alignments)  
4291.511 Million cell updates/sec

Title: US-09-849-729-1

Perfect score: 371  
Sequence: 1 gatcmgaaacgtytsgctc.....ctctctaccgacgacgac 371

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 33363688 seqs, 1658189874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_NA\_Main:

1: /cgn2\_6/ptodata/2/pna/US092A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US092B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US092C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US092D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US092E\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US092F\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US092G\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US092H\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US092I\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US092J\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US092K\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US092L\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US092M\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US092N\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US092O\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092P\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US092Q\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US092R\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US092S\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US092T\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US092U\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US092V\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US092W\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US092X\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US092Y\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US092Z\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US093A\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US093C\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US093D\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US093E\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US093F\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US093G\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US093H\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US093I\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US093J\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US093K\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US093L\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US093M\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US093N\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US093O\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US093P\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US093Q\_COMB.seq.\*

44: /cgn2\_6/ptodata/2/pna/US100A\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US100B\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US100C\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US100D\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US100E\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US100F\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US100G\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US100H\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US100I\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US100J\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US100K\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US100L\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US100M\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US100N\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US100O\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US100P\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US100Q\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US100R\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US100S\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US100T\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US100U\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US100V\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US100W\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US100X\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US100Y\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US100Z\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US101A\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US101B\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US101C\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US101D\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US101E\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US101F\_COMB.seq.\*  
76: /cgn2\_6/ptodata/2/pna/US101G\_COMB.seq.\*  
77: /cgn2\_6/ptodata/2/pna/US101H\_COMB.seq.\*  
78: /cgn2\_6/ptodata/2/pna/US101I\_COMB.seq.\*  
79: /cgn2\_6/ptodata/2/pna/US101J\_COMB.seq.\*  
80: /cgn2\_6/ptodata/2/pna/US101K\_COMB.seq.\*  
81: /cgn2\_6/ptodata/2/pna/US101L\_COMB.seq.\*  
82: /cgn2\_6/ptodata/2/pna/US101M\_COMB.seq.\*  
83: /cgn2\_6/ptodata/2/pna/US101N\_COMB.seq.\*  
84: /cgn2\_6/ptodata/2/pna/US101O\_COMB.seq.\*  
85: /cgn2\_6/ptodata/2/pna/US101P\_COMB.seq.\*  
86: /cgn2\_6/ptodata/2/pna/US101Q\_COMB.seq.\*  
87: /cgn2\_6/ptodata/2/pna/US101R\_COMB.seq.\*  
88: /cgn2\_6/ptodata/2/pna/US101S\_COMB.seq.\*  
89: /cgn2\_6/ptodata/2/pna/US101T\_COMB.seq.\*  
90: /cgn2\_6/ptodata/2/pna/US101U\_COMB.seq.\*  
91: /cgn2\_6/ptodata/2/pna/US101V\_COMB.seq.\*  
92: /cgn2\_6/ptodata/2/pna/US101W\_COMB.seq.\*  
93: /cgn2\_6/ptodata/2/pna/US101X\_COMB.seq.\*  
94: /cgn2\_6/ptodata/2/pna/US101Y\_COMB.seq.\*  
95: /cgn2\_6/ptodata/2/pna/US101Z\_COMB.seq.\*  
96: /cgn2\_6/ptodata/2/pna/US102A\_COMB.seq.\*  
97: /cgn2\_6/ptodata/2/pna/US102B\_COMB.seq.\*  
98: /cgn2\_6/ptodata/2/pna/US102C\_COMB.seq.\*  
99: /cgn2\_6/ptodata/2/pna/US102D\_COMB.seq.\*  
100: /cgn2\_6/ptodata/2/pna/US102E\_COMB.seq.\*  
101: /cgn2\_6/ptodata/2/pna/US102F\_COMB.seq.\*  
102: /cgn2\_6/ptodata/2/pna/US102G\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query length	ID	Description
1	369	99.5	371	35	US-09-849-729-1
2	366.6	98.8	372	35	US-09-849-729-2
3	50.2	13.5	1931	18	US-09-250-359-2
4	47	12.7	597	44	US-10-020-338-1346

```
5 45.4 12.2 373 24 US-09-540-234-4340 Sequence 4340, Ap
6 45.4 12.2 373 28 US-09-654-617-444082 Sequence 444082,
7 45.4 12.2 373 30 US-09-684-016-444082 Sequence 444082,
8 45.4 12.2 373 66 US-60-128-437-4849 Sequence 4849, Ap
9 45 12.1 390 23 US-09-790-399-77 Sequence 7, Appli
10 45 12.1 422 25 US-09-594-596-7755 Sequence 7755, Ap
11 45 12.1 422 28 US-09-654-617-439911 Sequence 439911,
12 45 12.1 422 30 US-09-684-016-439911 Sequence 439911,
13 45 12.1 422 53 US-10-304-123A-7755 Sequence 7755, Ap
14 45 12.1 422 53 US-10-304-123A-7755 Sequence 7755, Ap
15 45 12.1 561 34 US-09-804-730-7395 Sequence 7395, Ap
16 45 12.1 561 72 US-60-189-657-7478 Sequence 7478, Ap
17 45 12.1 603 28 US-09-654-617-436474 Sequence 436474,
18 45 12.1 603 28 US-09-684-016-436474 Sequence 436474,
19 44.8 12.1 685 53 US-10-437-963-45473 Sequence 45473, A
20 44.8 12.1 9588 27 US-09-620-392-41549 Sequence 41549, A
21 44.8 12.1 9588 31 US-09-702-134-6722 Sequence 6722, Ap
22 44.8 12.1 9588 34 US-09-815-264-60830 Sequence 60830, A
23 44.2 11.9 360 38 US-09-902-540-7498 Sequence 7498, Ap
24 44.2 11.9 5040 38 US-09-902-540-727 Sequence 727, Ap
25 43.4 11.7 789 51 US-10-366-683-9465 Sequence 9465, Ap
26 43.4 11.7 789 51 US-10-366-683-9465 Sequence 9465, Ap
27 43.4 11.7 918 51 US-10-366-683-9337 Sequence 9337, Ap
28 43.4 11.7 918 52 US-10-419-128-9337 Sequence 9337, Ap
29 43.4 11.7 1440 51 US-10-366-683-9411 Sequence 9411, Ap
30 43.4 11.7 1440 52 US-10-419-128-9411 Sequence 9411, Ap
31 43.2 11.6 568 27 US-09-620-392-17907 Sequence 17907, A
32 43.2 11.6 568 31 US-09-702-134-51103 Sequence 51103, A
33 43.2 11.6 568 34 US-09-815-264-108964 Sequence 108964,
34 43.2 11.6 3489 1 PCT-US99-27508-1 Sequence 1, Appli
35 43.2 11.6 3489 37 PCT-US99-27508-1 Sequence 1, Appli
36 43.2 11.6 3489 37 US-09-894-273-1 Sequence 1, Appli
37 43.2 11.6 3489 47 US-10-194-046-1 Sequence 1, Appli
38 43.2 11.6 8215 27 US-09-620-392-36718 Sequence 36718, A
39 43.2 11.6 8215 27 US-09-620-392-36718 Sequence 36718, A
40 43.2 11.6 12752 10 US-08-592-9638-60 Sequence 60, Appli
41 43.2 11.6 12752 10 US-08-592-9638-60 Sequence 60, Appli
42 43.2 11.6 12752 20 US-09-363-327-60 Sequence 63, Appli
43 43.2 11.6 12752 20 US-09-363-327-60 Sequence 63, Appli
44 43.2 11.6 12753 10 US-08-592-9638-59 Sequence 59, Appli
45 43.2 11.6 12753 10 US-08-592-9638-62 Sequence 62, Appli
```

## ALIGNMENTS

```
RESULT 1
US-09-849-729-1
Sequence 1, Application US/09849729
```

```
GENERAL INFORMATION:
APPLICANT: Liu, Jen-Kuei
```

```
APPLICANT: Lewis, Samantha
```

```
APPLICANT: Batz, Hans-Georg
```

```
APPLICANT: Ramaswamy, Latha
```

```
APPLICANT: Bohenzky, Roy
```

```
APPLICANT: Lin, Yu-Huei
```

```
APPLICANT: Montiel, Janine
```

```
APPLICANT: Chen, Benjamin
```

```
TITLE OF INVENTION: Sentinel Virus II
```

```
FILE REFERENCE: RDID 0070
```

```
CURRENT APPLICATION NUMBER: US/09/849,729
```

```
CURRENT FILING DATE: 2001-05-04
```

```
PRIOR APPLICATION NUMBER: US 60/202271
```

```
PRIOR FILING DATE: 2000-05-05
```

```
NUMBER OF SEQ ID NOS: 5
```

```
SOFTWARE: PatentIn version 3.0
```

```
SEQ ID NO 1
```

```
LENGTH: 371
```

```
TYPE: DNA
```

```
ORGANISM: Sentinel Virus II
```

```
US-09-849-729-1
```

```
Query Match
```

```
99.5%; Score 369; DB 35; Length 371;
```

Best Local Similarity 100.0%; Pred. No. 9e-81; Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GATCMGGAACGTTTSGCTGCTGATGACAGAGCGGWTGAAGCGGAGCGGATTTGA 60
1 GATCMGGAACGTTTSGCTGCTGATGACAGAGCGGWTGAAGCGGAGCGGATTTGA 60
Db 1 GATCMGGAACGTTTSGCTGCTGATGACAGAGCGGWTGAAGCGGAGCGGATTTGA 60
QY 61 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120
61 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120
Db 61 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120
QY 121 CGAGAACCAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180
121 CGAGAACCAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180
Db 121 CGAGAACCAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180
QY 181 GGTATTGACACCTCCGACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 240
181 GGTATTGACACCTCCGACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 240
Db 181 GGTATTGACACCTCCGACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 240
QY 241 GCGGCTTGGACGACCTGCTTGGACGACCTGACGACGACGACGACGACGACGACGAC 300
241 GCGGCTTGGACGACCTGCTTGGACGACCTGACGACGACGACGACGACGACGACGAC 300
Db 241 GCGGCTTGGACGACCTGCTTGGACGACCTGACGACGACGACGACGACGACGACGAC 300
QY 301 TCCTGCTTGAAGTGAATTTCCGACCCGCTTGAATTTCTCTCTCTCTCTCTCTCTAC 360
301 TCCTGCTTGAAGTGAATTTCCGACCCGCTTGAATTTCTCTCTCTCTCTCTCTCTAC 360
Db 301 TCCTGCTTGAAGTGAATTTCCGACCCGCTTGAATTTCTCTCTCTCTCTCTCTCTAC 360
QY 361 CGACGACGATC 371
361 CGACGACGATC 371
Db 361 CGACGACGATC 371
```

RESULT 2

US-09-849-729-2/c

Sequence 2, Application US/09849729

GENERAL INFORMATION:

APPLICANT: Liu, Jen-Kuei

APPLICANT: Lewis, Samantha

APPLICANT: Batz, Hans-Georg

APPLICANT: Ramaswamy, Latha

APPLICANT: Bohenzky, Roy

APPLICANT: Lin, Yu-Huei

APPLICANT: Montiel, Janine

APPLICANT: Chen, Benjamin

TITLE OF INVENTION: Sentinel Virus II

FILE REFERENCE: RDID 0070

CURRENT APPLICATION NUMBER: US/09/849,729

CURRENT FILING DATE: 2001-05-04

PRIOR APPLICATION NUMBER: US 60/202271

PRIOR FILING DATE: 2000-05-05

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 372

TYPE: DNA

ORGANISM: Sentinel Virus II

FEATURE:

NAME/KEY: misc.feature

LOCATION: (372)..(372)

OTHER INFORMATION: unknown: can be a, t, g, c or there could be no nucleotide found

OTHER INFORMATION: in this position.

US-09-849-729-2

Query Match

98.8%; Score 366.6; DB 35; Length 372;

Best Local Similarity 99.5%; Pred. No. 3.5e-80;

Matches 369; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCMGGAACGTTTSGCTGCTGATGACAGAGCGGWTGAAGCGGAGCGGATTTGA 60

Db 371 GATCMGGAACGTTTSGCTGCTGATGACAGAGCGGWTGAAGCGGAGCGGATTTGA 312

QY 61 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120

Db 61 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120

```
Db 311 CGACGACGACGACATTGCGATGAAAGATGGGACCGCGACGCTCTTGGCGGGCGGAGCG 252
Qy 121 CGAAGACCAAGACGACGAGAGGACGCTTAACGCGCCATCCGTTCTCTTCTGAGCG 180
Db 251 CGAAGACCAAGACGACGAGAGGACGCTTAACGCGCCATCCGTTCTCTTCTGAGCG 192
Qy 181 GGTATTGACACCTCGCGATTGCTGATCTGAAAGTTCTGCTTGGACGAGCGCTGATTGAC 240
Db 191 GGTATTGACACCTCGCGATTGCTGATCTGAAAGTTCTGCTTGGACGAGCGCTGATTGAC 132
Qy 241 GCGCGCTGCTGCGACCTGCTTGGACGCTGCAACCGCGACGCGTGGACCATGTCAGTT 300
Db 131 GCGCGCTGCTGCGACCTGCTTGGACGCTGCAACCGCGACGCGTGGACCATGTCAGTT 72
Qy 301 TCTGCTTGAAGAGTGAATTTCCGACCCCGTTGACCTTCTCTCTCTCTCTAC 360
Db 71 TCTGCTTGAAGAGTGAATTTCCGACCCCGTTGACCTTCTCTCTCTCTCTCTAC 12
Qy 361 CGACGACGATC 371
Db 11 CGACGACGATC 1

RESULT 3
US-09-250-359-2/c
; Sequence 2, Application US/09250359
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Basam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/250,359
; EARLIER FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: 09/130,114
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
; US-09-250-359-2

Query Match 13.5%; Score 50.2; DB 18; Length 1931;
Best Local Similarity 62.2%; Pred. No. 0.056;
Matches 79; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 31 GAAGACGCGGWTGAAGCGGACGCGATTTGACGACGACGACGACGATTTGGAAGATGG 90
Db 929 GGAAGACGCGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGG 870
Qy 91 GACCGCYGACGCTCTTGGCGGGCGGACGCGGAAACCAAGACGACGAGGACGAGGACGT 150
Db 869 GACCGCYGACGCGGAGGAGGACGCGGAGGAGGACGAGGACGCGGAGGAGGAGGAGGAGCA 810
Qy 151 CTAACGG 157
Db 809 GACCGGG 803

RESULT 4
US-10-020-338-1346
; Sequence 1346, Application US/10020338
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Hammond-Kosack, Kim
; APPLICANT: Urban, Martin
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52272)B
; CURRENT APPLICATION NUMBER: US/10/020,338
```

```
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,575
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 25642
; SEQ ID NO 1346
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Trifolium aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3833-014-Q1-K6-H10
US-10-020-338-1346

Query Match 12.7%; Score 47; DB 44; Length 597;
Best Local Similarity 58.4%; Pred. No. 0.3;
Matches 80; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

Qy 21 GGTGATGCGAAGACGCGGWTGAAGCGGACGCGGATTTGACGACGACGACGACATTGGCA 80
Db 337 GGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACG 396
Qy 81 TGAAGATGAGACCGCGYACGCTTGGCGGGCGGAGCGGAGGACGAGGACGAGGACGAGGACG 140
Db 397 AAGACGAGGACGCTCCATGAGGACGATGACGACGAGGAGGACGAGGACGAGGACGAGGACG 456
Qy 141 ACGAGGACGCTTACGCG 157
Db 457 ACGAGGACGACGACGAG 473

RESULT 5
US-09-540-234-4340
; Sequence 4340, Application US/09540234
; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy W.
; APPLICANT: Masucci, James D.
; TITLE OF INVENTION: Nucleic acid molecules and other molecules associated with
; FILE REFERENCE: 38-21(15726)B
; CURRENT APPLICATION NUMBER: US/09/540,234
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11840
; SEQ ID NO 4340
; LENGTH: 373
; TYPE: DNA
; ORGANISM: Trifolium aestivum
; OTHER INFORMATION: Clone ID: LIB103-056-Q1-B1-B1
US-09-540-234-4340

Query Match 12.2%; Score 45.4; DB 24; Length 373;
Best Local Similarity 57.7%; Pred. No. 0.72;
Matches 79; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

Qy 21 GGTGATGCGAAGACGCGGWTGAAGCGGACGCGGATTTGACGACGACGACGACATTGGCA 80
Db 176 GGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACG 235
Qy 81 TGAAGATGAGACCGCGYACGCTTGGCGGGCGGAGCGGAGGACGAGGACGAGGACGAGGACG 140
Db 236 AAGACGAGGACGCTCCATGAGGACGATGACGACGAGGAGGACGAGGACGAGGACGAGGACG 295
Qy 141 ACGAGGACGCTTACGCG 157
Db 296 ACTAGGACGACGACGAG 312

RESULT 6
US-09-654-617-444082
; Sequence 444082, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
```

;; CURRENT APPLICATION NUMBER: US/09/654,617  
;; CURRENT FILING DATE: 2000-09-05  
;; NUMBER OF SEQ ID NOS: 463173  
;; SEQ ID NO 444082  
;; LENGTH: 373  
;; TYPE: DNA  
;; ORGANISM: Triticum aestivum  
US-09-654-617-444082

Query Match 12.2%; Score 45.4; DB 28; Length 373;  
Best Local Similarity 57.7%; Pred. No. 0.72;  
Matches 79; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

QY 21 GGTGCATGCAAGAGCGGGTGAAGCGCGGATTGACGACGACGACATTGGCA 80  
DB 176 GGGACGAGAGCGAGAGCGAAGACTCCGACGAGGCGCTACGACGACGACGAAAGC 235  
QY 81 TGAAGATGGAGCCGCTGCTTGGCGGGCGGACGCGAGAACCAAGACGACGAG 140  
DB 236 AAGACGAGAGCTCCGATGAAAGCATGACCAAGAGAGAGAGAGAGAGAGAGAG 295  
QY 141 ACGAGGACGTCTACGCG 157  
DB 296 ACTAGACGACGACGAG 312

RESULT 7  
US-09-684-016-444082  
;; Sequence 444082, Application US/09684016  
;; GENERAL INFORMATION:  
;; APPLICANT: Liu, Jindong  
;; TITLE OF INVENTION: Annotated Plant Genes  
;; FILE REFERENCE: 38-21(15097)D  
;; CURRENT APPLICATION NUMBER: US/09/684,016  
;; CURRENT FILING DATE: 2000-10-10  
;; PRIOR APPLICATION NUMBER: 2000-09-05  
;; PRIOR FILING DATE: 2000-09-05  
;; NUMBER OF SEQ ID NOS: 463173  
;; SEQ ID NO 444082  
;; LENGTH: 373  
;; TYPE: DNA  
;; ORGANISM: Triticum aestivum  
US-09-684-016-444082

Query Match 12.2%; Score 45.4; DB 30; Length 373;  
Best Local Similarity 57.7%; Pred. No. 0.72;  
Matches 79; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

QY 21 GGTGCATGCAAGAGCGGGTGAAGCGCGGATTGACGACGACGACATTGGCA 80  
DB 176 GGGACGAGAGCGAGAGCGAAGACTCCGACGAGGCGCTACGACGACGACGAAAGC 235  
QY 81 TGAAGATGGAGCCGCTGCTTGGCGGGCGGACGCGAGAACCAAGACGACGAG 140  
DB 236 AAGACGAGAGCTCCGATGAAAGCATGACCAAGAGAGAGAGAGAGAGAGAGAG 295  
QY 141 ACGAGGACGTCTACGCG 157  
DB 296 ACTAGACGACGACGAG 312

RESULT 8  
US-60-128-437-4849  
;; Sequence 4849, Application US/60128437  
;; GENERAL INFORMATION:  
;; APPLICANT: Comer, Timothy W.  
;; APPLICANT: Masucci, James D.  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
;; FILE REFERENCE: 38-21(15725)A  
;; CURRENT APPLICATION NUMBER: US/60/128,437  
;; CURRENT FILING DATE: 1999-04-06

;; NUMBER OF SEQ ID NOS: 5533  
;; SEQ ID NO 4849  
;; LENGTH: 373  
;; TYPE: DNA  
;; ORGANISM: Triticum aestivum  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: LIB103-056-Q1-E1-B1  
US-60-128-437-4849

Query Match 12.2%; Score 45.4; DB 66; Length 373;  
Best Local Similarity 57.7%; Pred. No. 0.72;  
Matches 79; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

QY 21 GGTGCATGCAAGAGCGGGTGAAGCGCGGATTGACGACGACGACATTGGCA 80  
DB 176 GGGACGAGAGCGAGAGCGAAGACTCCGACGAGGCGCTACGACGACGACGAAAGC 235  
QY 81 TGAAGATGGAGCCGCTGCTTGGCGGGCGGACGCGAGAACCAAGACGACGAG 140  
DB 236 AAGACGAGAGCTCCGATGAAAGCATGACCAAGAGAGAGAGAGAGAGAGAGAG 295  
QY 141 ACGAGGACGTCTACGCG 157  
DB 296 ACTAGACGACGACGAG 312

RESULT 9  
US-09-790-399-7  
;; Sequence 7, Application US/09790399  
;; GENERAL INFORMATION:  
;; APPLICANT: Gold, Larry  
;; APPLICANT: Tuerk, Craig  
;; APPLICANT: Pribnow, David  
;; APPLICANT: Smith, Jonathan D.  
;; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation  
;; FILE REFERENCE: NEX02/CI-CO2  
;; CURRENT APPLICATION NUMBER: US/09/790,399  
;; CURRENT FILING DATE: 2001-02-22  
;; PRIOR APPLICATION NUMBER: 09/197,649  
;; PRIOR FILING DATE: 1998-11-23  
;; PRIOR APPLICATION NUMBER: 07/829,461  
;; PRIOR FILING DATE: 1992-01-31  
;; PRIOR APPLICATION NUMBER: 07/739,055  
;; PRIOR FILING DATE: 1991-08-01  
;; PRIOR APPLICATION NUMBER: 07/561,968  
;; PRIOR FILING DATE: 1990-08-02  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO 7  
;; LENGTH: 390  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Sequence  
;; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed  
;; OTHER INFORMATION: fragments having NcoI restriction sites.  
US-09-790-399-7

Query Match 12.1%; Score 45; DB 33; Length 390;  
Best Local Similarity 57.8%; Pred. No. 0.9;  
Matches 78; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 21 GGTGCATGCAAGAGCGGGTGAAGCGCGGATTGACGACGACGACATTGGCA 80  
DB 1 GGGCCATGAG 60  
QY 81 TGAAGATGGAGCCGCTGCTTGGCGGGCGGACGCGAGAACCAAGACGACGAG 140  
DB 61 ACGACGAG 120  
QY 141 ACGAGGACGTCTACG 155  
DB 121 ACGACGAGAGAGAG 135



```

Db      75  GCTCAAGAAATGGCTCAACTCGAGCAAGACGCCAAGTTCAACGCCGACGAGACGA 134
Qy      61  CGACGACGACGACATTGTCGATGAAAGATGGGACCGCTGACCTTGGCGGGCGGAGCG 120
Db      135  CGGCGACGACGACGCGGAGCAAGAGAGAGCTGCGCTGCGACCGCGCCGAGCGGAGCG 194
Qy      121  CGAAGACCAAGACGACGAGACGAGAGCG 149
Db      195  CGAGGCCGAGGCGTGGACATCGCGGAGG 223

```

```

RESULT 14
US-10-626-717-7755
; Sequence 7755, Application US/10626717
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: De La Pena, Robert C.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15878) D
; CURRENT APPLICATION NUMBER: US/10/626,717
; PRIOR FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 10/304,123
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: US 09/594,596
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 10952
; SEQ ID NO 7755
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(422)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: uc-tbrow189025e10b1
; US-10-626-717-7755

```

```

Query Match      12.1%; Score 45; DB 53; Length 422;
Best Local Similarity 54.4%; Pred. No. 0.91;
Matches 81; Conservative 4; Mismatches 64; Indels 0; Gaps 0;

Qy      1  GATCWGAAGACGTTTSGCTCGTGCTATGCAAGGACCGGWTGAAGCGGACGGATTGA 60
Db      75  GCTCAAGAAATGGCTCAACTCGAGCAAGACGCCAAGTTCAACGCCGACGAGACGA 134
Qy      61  CGACGACGACGACATTGTCGATGAAAGATGGGACCGCTGACCTTGGCGGGCGGAGCG 120
Db      135  CGGCGACGACGACGCGGAGCAAGAGAGAGCTGCGCTGCGACCGCGCCGAGCGGAGCG 194
Qy      121  CGAAGACCAAGACGACGAGACGAGAGCG 149
Db      195  CGAGGCCGAGGCGTGGACATCGCGGAGG 223

```

```

RESULT 15
US-09-804-730-7395
; Sequence 7395, Application US/09804730
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Hammond-Kosack, Kim
; APPLICANT: Masucci, James D.
; APPLICANT: Urban, Martin
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51838) B
; CURRENT APPLICATION NUMBER: US/09/804,730
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: US 60/189,657
; PRIOR FILING DATE: 2000-03-15

```

```

; NUMBER OF SEQ ID NOS: 22828
; SEQ ID NO 7395
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(561)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3352-008-P1-K1-H2
; US-09-804-730-7395

```

```

Query Match      12.1%; Score 45; DB 34; Length 561;
Best Local Similarity 56.6%; Pred. No. 0.94;
Matches 81; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

Qy      19  TCGGTGCATGCAAGAGACCGGWTGAAGCGGATTTGACGACGACGACGACATTGC 78
Db      3  TCCGCGAGCGCTGGGGACGACGACCAAGAGAGACCCCGACATGACGACGACGACGA 62
Qy      79  GATGAAGATGGGACCGCTGACCTTGGCGGGCGGAGCGCGAAGACCAAGACGACGA 138
Db      63  GGAGGACCCCGACCAAGACGACGACCAATATGACGACGACGACGACGACGACGA 122
Qy      139  GGAAGAGACGCTCTTACGCGCGCA 161
Db      123  TGAAGACGACGACCAAGACGACGA 145

```

```

Search completed: September 15, 2003, 01:23:00
Job time : 2870 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2003, 23:24:09 ; Search time 75 Seconds  
(without alignments)  
4067.492 Million cell updates/sec

Title: US-09-849-729-1

Sequence: 1 gatcmggaacgyttsctc.....ctctctaccgcgacgcgacgc 371

Scoring table: IDENTITY\_NUC

Searched: 601805 seqs, 411134613 residues

Total number of hits satisfying chosen parameters: 1203610

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```
Database : Pending Patents NA.New:*
1 : /sgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2 : /sgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3 : /sgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4 : /sgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5 : /sgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6 : /sgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7 : /sgn2_6/ptodata/2/pna/US16_NEW_COMB.seq:*
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
C	1	40.8	11.0	1827	6	US-10-425-1144-21585	Sequence 21585, A
	2	40.4	10.9	1099	6	US-10-425-1144-13792	Sequence 13792, A
	3	39.6	10.7	15559	6	US-10-646-664-1	Sequence 1, Appl1
	4	39.4	10.6	3435	7	US-60-485-450-67	Sequence 67, Appl1
	5	39.4	10.6	100537	7	US-60-485-450-11874	Sequence 11874, A
C	6	38.8	10.5	993	6	US-10-425-1144-34768	Sequence 34768, A
	7	38.8	10.5	1068	6	US-10-425-1144-34686	Sequence 34686, A
	8	38.2	10.5	864	6	US-10-425-1144-19765	Sequence 19765, A
	9	38.2	10.3	950	6	US-10-425-1144-6704	Sequence 6704, Ap
	10	38.2	10.3	965	6	US-10-425-1144-19745	Sequence 19745, A
C	11	38.2	10.3	1046	6	US-10-425-1144-35806	Sequence 35806, A
	12	38.2	10.3	1366	6	US-10-425-1144-5274	Sequence 5274, Ap
	13	38.2	10.3	1451	6	US-10-425-1144-5525	Sequence 5525, Ap
	14	38.2	10.3	1498	6	US-10-425-1144-5524	Sequence 5524, Ap
	15	38.2	10.3	1633	6	US-10-425-1144-11354	Sequence 11354, A
C	16	38.2	10.3	1882	6	US-10-425-1144-31627	Sequence 31627, A
	17	38.2	10.3	2109	6	US-10-425-1144-33989	Sequence 33989, A
	18	38.2	10.3	2169	6	US-10-425-1144-33293	Sequence 33293, A
	19	38.2	10.3	2173	6	US-10-425-1144-5896	Sequence 5896, Ap
	20	38.2	10.3	2179	6	US-10-425-1144-5774	Sequence 5774, Ap
C	21	38.2	10.3	2182	6	US-10-425-1144-25314	Sequence 25314, Ap
	22	38.2	10.3	2197	6	US-10-425-1144-20309	Sequence 20309, A
	23	38.2	10.3	2199	6	US-10-425-1144-5526	Sequence 5526, Ap
	24	38.2	10.3	2200	6	US-10-425-1144-32116	Sequence 32116, A
	25	38.2	10.3	2205	6	US-10-425-1144-5793	Sequence 5793, Ap
26	38.2	10.3	2207	6	US-10-425-1144-22284	Sequence 22284, A	

## ALIGNMENTS

27	38.2	10.3	2309	6	US-10-42S-114A-34008	Sequence 34008, A
28	38.2	10.3	2210	6	US-10-42S-114A-5784	Sequence 5784, Ap
29	38.2	10.3	2211	6	US-10-42S-114A-1166	Sequence 1166, Ap
30	38.2	10.3	2211	6	US-10-42S-114A-7049	Sequence 7049, Ap
31	38.2	10.3	2213	6	US-10-42S-114A-5775	Sequence 5775, Ap
32	38.2	10.3	2213	6	US-10-42S-114A-22689	Sequence 22689, A
33	38.2	10.3	2213	6	US-10-42S-114A-32749	Sequence 32749, A
34	38.2	10.3	2214	6	US-10-42S-114A-25552	Sequence 25552, A
35	38.2	10.3	2215	6	US-10-42S-114A-5781	Sequence 5781, Ap
36	38.2	10.3	2218	6	US-10-42S-114A-33345	Sequence 33345, A
37	38.2	10.3	2648	6	US-10-42S-114A-26406	Sequence 26406, A
38	38.2	10.1	1985	6	US-10-42S-114A-17820	Sequence 17820, A
39	37.6	10.1	1627	6	US-10-42S-114A-25600	Sequence 25600, A
40	37.6	10.1	1696	6	US-10-42S-114A-22836	Sequence 22836, A
41	37.6	10.1	1866	6	US-10-42S-114A-897	Sequence 897, App
42	37.6	10.1	710	6	US-10-42S-114A-234	Sequence 234, App
43	37.4	10.1	719	6	US-10-42S-114A-2518	Sequence 2518, App
44	37.4	10.1	1312	6	US-10-42S-114A-20197	Sequence 20197, A
45	37.4	10.1	1557	6	US-10-42S-114A-28931	Sequence 28931, A

```

RESULT 1
US-10-425-114A-21585/c
; Sequence 21585, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 21585
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-032-G5_FLI
US-10-425-114A-21585

Query Match      11.0% Score 40.8; DB 6; Length 1827;
Best Local Similarity 51.9%; Pred. No. 0.019;
Matches 84; Conservative 3; Mismatches 75; Indels 0; Gaps 0;

Qy      3  TCMGGAACGATTGTCGTCGATGACGAAGAAGACGGGTGAAGCGGACGGATTGACG 62
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      618  TCCTGGAAAGCGCGCCGGGACACGGCGCTCCAGTTCGGCAGGTGAGAGTGAAGCTGAGC 559

Qy      63  ACGACGACGACATTGCGATGTAAGATGGAGACCGCYAGACGTCCTTGGCGGGGCGAGCGCG 122
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      558  ACGTGCAGGTTGATGTGGAGAAGATGCACCTGCGGGCGGACCCCGCGCGGCGCGCG 499

Qy      123  AGAACCAAGACGACGAGGACGACGACGCTCTACGCGCGCATCC 164
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      498  TCCAGCGAGTGGCGGCCGGAAGAGTGGCCACGAACGCGCGC 457

RESULT 2
US-10-425-114A-13792/c
; Sequence 13792, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E

```

```

; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13792
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: LIB143-044-G12_FLI
US-10-425-114A-13792

Query Match          10.9%; Score 40.4; DB 6; Length 1099;
Best Local Similarity 53.2%; Pred. No. 0.021;
Matches 83; Conservative 1; Mismatches 72; Indels 0; Gaps 0;

QY 34 GAGCGGWTGAAGCGGACGGGATTGACGACGACGACGACGATTCGATGAAAGATGGAC 93
DB 648 GAGCGCGATGAGGACCGCGCGCGGAGGTCGCGGGAAGCTGGCGGAAGAGCGCGC 589
QY 94 CGCGAGGTCCTTGGCGGGGCGGAGCGCGGAACCAAGACGACGAGGACGAGCTCTA 153
DB 588 CGGCGAGGACGCGCGGATGAGAGGCGGCGGACGCGGAGGAGGAGCGGTGTAGTAC 529
QY 154 CGCGCGCATCCGTTCTTCTTCTGAGCGGATTTGA 189
DB 528 CTTCCAGCCCCGCGGTGTGTGTGAGTGAAGTAGA 493

RESULT 3
US-10-646-664-1
; Sequence 1, Application US/1064664
; GENERAL INFORMATION:
; APPLICANT: Shen, Ben
; APPLICANT: Hyung-Jin, Kwon
; TITLE OF INVENTION: METHODS OF DIRECTING C-O BOND FORMATION UTILIZING A TYPE II
; TITLE OF INVENTION: POLYKETIDE SYNTHASE SYSTEM
; FILE REFERENCE: 054030-0031
; CURRENT APPLICATION NUMBER: US/10/646,664
; CURRENT FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: US 60/405,245
; PRIOR FILING DATE: 2002-08-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 15559
; TYPE: DNA
; ORGANISM: Streptomyces griseus
US-10-646-664-1

Query Match          10.7%; Score 39.6; DB 6; Length 15559;
Best Local Similarity 49.5%; Pred. No. 0.086;
Matches 99; Conservative 1; Mismatches 100; Indels 0; Gaps 0;

QY 18 CTGCGTGCATGCAGGAAGACGGGWTGAAGCGGACGGGATTTGACGACGACGACATTG 77
DB 15095 CGCTGACCTGGACGATCTGGGTCTGACGCGGAGAGGTCGTGCGGGTGGTGA 15154
QY 78 CGATGAAGATGAGACCCGCGCGCTCTTGGGGGGCGGAGCGGAGAACCAAGACGACG 137
DB 15155 CGGTGGATTCGCCGACCGAGCGTCCGGCGCGGACGCGGCGCATCTGTAAGACG 15214
QY 138 AGGACGAGACGCTTACGCGCGCATCCGTTCTTCTGAGCGGGTATTTGACACCTCG 197
DB 15215 AGGCGAGGCGGCGGACGAGCTCGCGGGTCTCTGCGGCGGACGAAGTTCACTAGTCG 15274
QY 198 CATTGCTGATCTCTGAATTTC 217
DB 15275 CATGTCTTCCCGCTGTC 15294
```

```

RESULT 4
US-60-485-450-67
; Sequence 67, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-Yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/485,450
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 3435
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-485-450-67

Query Match          10.6%; Score 39.4; DB 7; Length 3435;
Best Local Similarity 49.8%; Pred. No. 0.06;
Matches 100; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 28 GCAGAGGACCGGWTGAAGCGGACGGGATTGACGACGACGACGACATTGCGATGAAGA 87
DB 864 GAGAGGAGAGAGAGAGAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGA 923
QY 88 TGGACCGCGYACGTCCTTGGCGGGCGGAGCGCGGAACCAAGACGACGAGACGAGA 147
DB 924 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 983
QY 148 CGTCTACGCGCGCATCCGTTCTTCTTCTGAGCGGGATTTGACACTCCGATTCGTAT 207
DB 984 GGAGGACGCTCCGATGTGATCTTTCAGAGAGACACTCTGCACCTTGGCCAGAGGC 1043
QY 208 CCTGAAGTCTCGCTTGACAGA 228
DB 1044 CCTGAGCTCCGGGCCCCAGA 1064

RESULT 5
US-60-485-450-11874
; Sequence 11874, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-Yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/485,450
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11874
; LENGTH: 100537
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-485-450-11874

Query Match          10.6%; Score 39.4; DB 7; Length 100537;
Best Local Similarity 49.8%; Pred. No. 0.18;
Matches 100; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 28 GCAGAGGACCGGWTGAAGCGGACGGGATTGACGACGACGACGACATTGCGATGAAGA 87
DB 61771 GAGAGAGAGAGAGAGAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGA 61830
```





```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 6704  
; LENGTH: 950  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700575728_FLI  
US-10-425-114A-6704
```

```
Query Match 10.3%; Score 38.2; DB 6; Length 950;  
Best Local Similarity 54.1%; Pred. No. 0.09;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;
```

```
QY 28 GCAGAAAGACGGCGGTGAAGCGCGGATTGACGACGACGACATTGGCATGAAGA 87  
DB 346 GAAGAAGACGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
QY 88 TGGGACCGCYGACGCTCTTGGCGGGGCGAGCGCGAGAACCAAGACGACGAGAG 147  
DB 406 GAAGGACGGCGGCGGCGAGCGGCGGAGAACCAAGAGAGAGCGGCGGAGAGAG 465  
QY 148 CGGTACGCGCGCAT 162  
DB 466 CCGGCGCGCGCAT 480
```

## RESULT 10

```
US-10-425-114A-19745  
; Sequence 19745, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 19745  
; LENGTH: 965  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3151-008-H4_FLI  
US-10-425-114A-19745
```

```
Query Match 10.3%; Score 38.2; DB 6; Length 965;  
Best Local Similarity 54.1%; Pred. No. 0.09;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;
```

```
QY 28 GCAGAAAGACGGCGGTGAAGCGCGGATTGACGACGACGACATTGGCATGAAGA 87  
DB 477 GAAGAAGACGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536  
QY 88 TGGGACCGCYGACGCTCTTGGCGGGGCGAGCGCGAGAACCAAGACGACGAGAG 147  
DB 537 GAAGGACGGCGGCGGCGAGCGGCGGAGAACCAAGAGAGAGCGGCGGAGAGAG 596  
QY 148 CGGTACGCGCGCAT 162  
DB 597 CCGGCGCGCGCAT 611
```

## RESULT 11

```
US-10-425-114A-35806  
; Sequence 35806, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 35806  
; LENGTH: 1046  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mexicana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE002D07_FLI  
US-10-425-114A-35806
```

```
Query Match 10.3%; Score 38.2; DB 6; Length 1046;  
Best Local Similarity 54.1%; Pred. No. 0.093;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;
```

```
QY 20 CGGTGATGCAAGAGACGGCGGTGAAGCGCGGATTGACGACGACGACATTGGC 79  
DB 599 CGCAACACGATCAAGAGAGAGAGAGATCGGTCGCAAGTTGCTGCGCCACCAAGAG 658  
QY 80 ATGAAGATGGGACCGCGCYGACGCTTGGCGGGGCGAGCGGAGAACCAAGAGAG 139  
DB 659 ATGAAGAGCGCATCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 718  
QY 140 GACGAGACGCTTAC 154  
DB 719 GACGAGTTGATGAC 733
```

## RESULT 12

```
US-10-425-114A-5274  
; Sequence 5274, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 5274  
; LENGTH: 1366  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700452211_FLI  
US-10-425-114A-5274
```

```
Query Match 10.3%; Score 38.2; DB 6; Length 1366;  
Best Local Similarity 54.1%; Pred. No. 0.1;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;
```

```
QY 20 CGGTGATGCAAGAGACGGCGGTGAAGCGCGGATTGACGACGACGACATTGGC 79  
DB 920 CGCAACACGATCAAGAGAGAGAGATCGGTCGCAAGTTGCTGCGCCACCAAGAG 979  
QY 80 ATGAAGATGGGACCGCGCYGACGCTTGGCGGGGCGAGCGGAGAACCAAGAGAG 139
```

Db 980 ATGAGAGACCCATTCAGCGGCGCATCGCTGCTGAGCCCAACCAAGCTCGCCGAGGCG 1039  
QY 140 GACGAGACGCTTAC 154  
Db 1040 GACGAGTTGATGAC 1054

## RESULT 13

US-10-425-114A-5525  
; Sequence 5525, Application US/10425114A  
; GENERAL INFORMATION:

APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 5525  
LENGTH: 1431  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700464617\_FLI

US-10-425-114A-5525

Query Match 10.3%; Score 38.2; DB 6; Length 1431;  
Best Local Similarity 54.1%; Pred. No. 0.1;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

QY 20 CGGTGATGCAGAAAGACGGGWTGAAGCGGACGGATTGACGACGACGACGATTGCG 79

Db 984 CGCAACACGATCAAGAGCAGAAAGATCGGTCACAGCTGCTGCCGACGACGAGAGAG 1043

QY 80 ATGAAGATGGAGACCGCCYACGCTCTTGCGGGGCGGACGCGAGAACCAAGCAGCGAG 139

Db 1044 ATCGAGAGCGCCATTCAGCGGCGCATCGCTGCTGAGCCCAACGACTCGCCGAGGCG 1103

QY 140 GACGAGACGCTTAC 154

Db 1104 GACGAGTTGATGAC 1118

## RESULT 14

US-10-425-114A-5524  
; Sequence 5524, Application US/10425114A  
; GENERAL INFORMATION:

APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 5524  
LENGTH: 1498  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700458467\_FLI

US-10-425-114A-5524

Query Match 10.3%; Score 38.2; DB 6; Length 1498;  
Best Local Similarity 54.1%; Pred. No. 0.1;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

QY 20 CGGTGATGCAGAAAGACGGGWTGAAGCGGACGGATTGACGACGACGACGATTGCG 79

Db 1052 CGCAACACGATCAAGAGCAGAAAGATCGGTCACAGCTGCTGCCGACGACGAGAGAG 1111

QY 80 ATGAAGATGGAGACCGCCYACGCTCTTGCGGGGCGGACGCGAGAACCAAGCAGCGAG 139

Db 1112 ATCGAGAGCGCCATTCAGCGGCGCATCGCTGCTGAGCCCAACGACTCGCCGAGGCG 1171

QY 140 GACGAGACGCTTAC 154

Db 1172 GACGAGTTGATGAC 1186

## RESULT 15

US-10-425-114A-17354  
; Sequence 17354, Application US/10425114A  
; GENERAL INFORMATION:

APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 17354  
LENGTH: 1643  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3069-015-A8\_FLI

US-10-425-114A-17354

Query Match 10.3%; Score 38.2; DB 6; Length 1643;  
Best Local Similarity 54.1%; Pred. No. 0.11;  
Matches 73; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

QY 20 CGGTGATGCAGAAAGACGGGWTGAAGCGGACGGATTGACGACGACGACGATTGCG 79

Db 1223 CGCAACACGATCAAGAGCAGAAAGATCGGTCACAGCTGCTGCCGACGACGAGAGAG 1282

QY 80 ATGAAGATGGAGACCGCCYACGCTCTTGCGGGGCGGACGCGAGAACCAAGCAGCGAG 139

Db 1283 ATCGAGAGCGCCATTCAGCGGCGCATCGCTGCTGAGCCCAACGACTCGCCGAGGCG 1342

QY 140 GACGAGACGCTTAC 154

Db 1343 GACGAGTTGATGAC 1357

Search completed: September 15, 2003, 01:24:35  
Job time : 76 secs

**THIS PAGE BLANK (USPTO)**